194-00313TPW

ECS User Characterization Methodology and Results

White Paper
Working Paper

Technical Paper—Not intended for formal review or government approval.

September 1994

Prepared Under Contract NAS5-60000

RESPONSIBLE SCIENTIST/ENGINEER

Lori J. Tyahla /s/ Sept. 30, 1994

Lori J. Tyahla, Science Specialist Date
EOSDIS Core System Project

SUBMITTED BY

Robert J. Curran /s/ Sept. 30, 1994

Robert Curran, Science Office Date
EOSDIS Core System Project

Hughes Applied Information Systems, Inc.
Landover, Maryland

This page intentionally left blank.

Contents

1.0 Introduction

2.0 User Scenario Matrix Purpose and Development

3.0 Scenario Collection and Analysis

4.0 ECS User Demographics

5.0 System Access Characteristics

6.0 Data Access Characteristics

Figures

Tables

Appendix A

Abbreviations and Acronyms

1.0 Introduction

1.1 Purpose

In order to design a data and information system that is responsive to the needs of users, it is advantageous to first estimate the size and needs of the user community. This knowledge can then be used to model system performance under varying load conditions. For example, a performance model can answer the question "If 1,000 users are simultaneously accessing the system, how long will it take for the system to send a browse image to a specific user?". The user characteristics can also be used to determine how users intend to access and/or subset data; for example, do the majority of users require data subsetting and, if so, what types of subsetting will they require (spatial, temporal, spectral, parametric)? Realistic information regarding users' data access patterns will enable ECS developers to design a system that is both user-friendly and highly efficient. This is the goal of the ECS modeling activity.

The ECS modeling activity is composed of several sub-models. There are three major component models: the User Model, the Data Model, and the Performance Model. In addition, the Performance Model is further subdivided into the Activation Model, the Static Model, the Algorithm Model, the Dynamic Model, the Quasi-Dynamic Model, and a Technology database. It is not the intent of this document to describe the overall modeling process; the focus here is on the goals and methodologies of the User Model.

The purpose of this Technical Paper is to document the methodologies used to characterize the ECS User community as of the second quarter of 1994. In addition, this document details the interfaces between the User Characterization Team and several other design groups, including both developers and modelers. Because the ECS user community is very large and diverse, several methods were employed to describe both the demographics of the community as well as the interactions of the user community with EOSDIS. This document explains the origin and subsequent development of the User Scenario Matrix, the processes of scenario collection and analysis, and the methods used to obtain demographic information about the ECS user community.

A high-level summary of the results of the user characterization effort can be found in the document, *User Characterization and Requirement Analysis* (doc # 194-00312TPW) and the reader is encouraged to obtain that document as it contains results which are not necessarily part of this document. Other relevant documents include: *ECS User/Data Model Approach and Plan* (June, 1993) and *User Scenario Notebook* (doc # 194-00311TPW).

1.2 Organization

In general, this document is organized chronologically beginning with the development of the user scenario methodology in early 1993 and continuing to the results presented at the Preliminary Design Review (PDR) in June of 1994. Section 2 details the development of the scenario-based methodology. Section 3 describes the scenario collection and analysis process.

Section 4 details the methodologies and sources employed to estimate the demographics of the EOSDIS user community. Section 5 describes system access patterns of the user community and Section 6 describes the likely data access patterns of ECS users.

1.2.1 User Scenario Matrix Purpose and Development

Section 2, User Scenario Matrix Purpose and Development, describes the purpose of the User Scenario Matrix and its development. In section 2.2, three previous user classification schemes (Barkstrom's 1991 model, the 1993 NASA Ad Hoc Working Group model , and the Hughes Team "Hats" model) are described briefly. Section 2.3 details the overall approach to the selection of scenarios. Section 2.4 tracks the development of the current User Scenario Matrix, from the beginning (based on the three previous approaches) to revisions made in January, 1994. Included in this section are the current matrix itself, the current definitions of the user classes (rows and columns of matrix), and two decision trees used to place users into a matrix cell.

1.2.2 Scenario Collection and Analysis

Section 3 describes the scenario collection and analysis process. Section 3.2 details the scenario collection methods. Section 3.3 describes how the scenario information was translated into engineering data for the purpose of building service threads. Section 3.4 describes the manner in which the user scenarios were analyzed for functional requirements. The validation of the scenario matrix can be found in Section 3.5. Section 3.6 describes the results of the review of the science scenarios by the science user working groups. Section 3.7 explains how the scenario information was input to other modeling activities.

1.2.3 EOSDIS User Demographics

Section 4 describes the methods used to determine the demographics of the ECS user community. Three high-level communities are defined in Section 4.1: EOS Science users, General Science Users, and Non-Science users. The methods to determine the demographics of the EOS Science community are outlined in Section 4.2, followed by the methods employed to determine the demographics of the General Science community (Section 4.3) and the Non-Science user community (Section 4.4). Section 4.5 contains a summary of the demographics of the individual user communities.

1.2.4 System Access Characteristics

Section 5 begins with an introduction (Section 5.1) that defines terms such as access frequency, access method, and access path. Following the introduction are Section 5.2 and 5.3 that provide methodologies employed and the resulting system characteristics for the General Science community and the Non-Science community, respectively.

1.2.5 Data Access Characteristics

Section 6 contains the data access characteristics of the user community. Section 6.1 explains the importance of data access characteristics. Section 6.2 discusses interest in EOS data by layer in the data pyramid and Section 6.3 discusses data access characteristics from a volumetric point-of-view; specifically the data volumes staged and distributed to users.

1.3 Review and Approval

This White Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

The ideas expressed in this White Paper are valid for the period of time beginning in April 1994 until PDR, at which time additional methodologies may be in use; the concepts presented here are expected to migrate into the following formal CDRL deliveries:

Questions regarding technical information contained within this Paper should be addressed to the following ECS contacts:

- ECS Contacts
  - Lori J. Tyahla, Science Specialist, ESSi, (301) 925-0803, ltyahla@eos.hitc.com
  - Pitt Thome, ESSi, (301) 925-0807, pthome@eos.hitc.com
  - Celeste Jarvis, Program Manager, ESSi, (301) 925-0800, cjarvis@eos.hitc.com

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems, Inc.
1616A McCormick Dr.
Landover, MD 20785

2. User Scenario Matrix Purpose and Development

2.1 Introduction

A critical aspect of characterizing the ECS user community is interaction with members of the community to obtain detailed information on how they intend to use EOSDIS. One of the mechanisms by which this information exchange takes place is through the development of user scenarios. Because the expected user community is so large, it is not feasible to speak with each and every user; thus the number of scenarios must be limited to a manageable size. However, it is very important that the sample scenarios be representative of the entire user community. In other words, there can be no biases with respect to research discipline, geographic scale of research, or any other type of user attribute.

The purpose of this section is to outline the development of the scenario-based methodology used to characterize the ECS user community. The methodology is based on work from three sources: the NASA User Model / Data Model Ad Hoc Working Group, Bruce Barkstrom, and the Hughes Team; section 2.2 of this paper summarizes the work conducted by these groups as it pertains to user classification schemes. Section 2.3 discusses the overall approach used in identifying an appropriate classification scheme and presents the resulting User Matrix based on this work. Section 2.4 discusses the refocusing of the User Matrix in January of 1994 and Section 2.5 describes methods for categorizing the widely diverse user community into the matrix itself.

2.2 Previous User Classification Schemes

A great deal of work has been done by several groups and individuals with regard to categorizing the ECS user community. Three major efforts were performed and documented by the NASA Ad Hoc Working Group (Ad Hoc Working Group, 1993), Bruce Barkstrom (Barkstrom, 1991), and the Hughes Team (Hughes Team, internal document, 1991). Each of these efforts and its approach to categorizing the user community is described briefly below. It should be noted that regardless of the approach taken, each group recognizes that an individual user is likely to "belong" to more than one user category. The approach adopted by the ECS User Characterization Team is a combination of all three of the methods described below.

2.2.1 Barkstrom's User Classification

Barkstrom's User Model addresses three issues: size of the user community, user interaction with EOSDIS, and data volume required by researchers. His model also focuses primarily on Earth Science research community; though he does make some suggestions regarding ECS use by the non-research community. Three methods are presented to determine the size of the Earth Science research community; the first estimates the size of the EOS Investigator user group by counting the Principal Investigators, Co-Investigators, Facility Team Members. Added to this count is an average number of support personnel per investigation. The second and third methods estimate the number of Earth Science researchers by counting publications and by analyzing user requests for digital data from various archives, respectively (Barkstrom, 1991). In addition, Barkstrom also estimates the growth in the user community based on demographic considerations as well as on the growth in funding of scientific endeavors.

The issue of user interaction with EOSDIS is addressed by classifying user research activities into five categories: Case studies, Field studies, Climatologies, Theoretical studies, and Reviews. Barkstrom deduces the fraction of scientific work in each category by performing a survey of publications in *Journal of Geophysical Research, Journal of Glaciology* , and *Bioscience*. He also presents four categories of researcher collaboration (single author, single institution; several authors, single institution; few (2-3) authors, several institutions; and several (4 or more) authors, several institutions) and estimates the frequency of collaboration types for each research activity.

2.2.2 NASA Ad Hoc Working Group User Classification

The work performed by the NASA Ad Hoc Working Group focused on the Earth Science community. These users are initially categorized with respect to System Access patterns ranging from "Person-to-Person" to "High Tech". A user in the "Person-to-Person" category prefers to use the telephone to obtain information on available data, while a "High Tech" user has high tech workstations and "intelligent" software and is connected to various networks. In all, there are 17 representative user classes. These classes are then consolidated to form a set of 7 classes of user system access needs, Person-to-Person, Conservative Tech, Intermediate, Advanced (Browse), Advanced (Analysis), Advanced (Production), and High Tech.

In addition, the Ad Hoc Working Group categorized the users according to four broad types of activities: Site (Local) Studies, Regional Studies, Global Studies, and Instrument/Calibration Studies. A matrix was then formed of Data Access vs. System Access. Each cell in this matrix has a representative user scenario and a "resource envelope" derived from that scenario (McConaughy et al., 1993).

2.2.3 The Hughes Team "Hats" User Classification

The Hughes Team "Hats" Model approached the problem of determining how the users will access EOSDIS and what services they will require by characterizing the "types of tasks" the users perform. The Hughes Team identified 16 "hats" that a given individual may wear; some examples of these hats are Modeler, Hypothesis Tester, and Instrument Builder. Again, it must be emphasized that an individual may wear more than one hat simultaneously or may change hats over time. A researcher will perform certain tasks while wearing a particular hat (which may or may not be unique to that hat) and the services and data associated with each task are identified. Some of these services and data are provided by the EOSDIS and some are not.

2.3 Overall Approach to Scenario Selection

Given the three approaches outlined above, our goal was to adopt the strengths found in these approaches as the basis for a new approach to meet current system modeling needs. The central objective in the current approach is to select a set of scenarios that represents the entire ECS user community. In order to do this, the set of scenarios selected must meet several criteria:

1. The selected set must capture the variability found in the user community
2. The user classes must be "countable" for modeling purposes.
3. The number of scenarios required should be limited to what can realistically be developed within time and manpower constraints.

Given these criteria, a conceptual approach was developed based on the concepts used in principal components analysis. The user community and system usage can be viewed from many points of view, as is indicated in the discussion above. The limiting factor is the number of scenarios that can realistically be supported. At one end of the spectrum, a scenario could be developed for each user, on the other end, one scenario could be developed to represent all users. A more moderate approach would be to develop an n-dimensional matrix with various user characteristics (system access patterns, discipline, work habits etc.) listed on each axis. Though this approach is likely to capture the variability in the user community, the number of scenarios required to support this would still be unrealistically large.

Using the same concepts used in principal components analysis, the ideal matrix would provide a view that shows the maximum variability in the user community along two dimensions. The difficulty lies in conceptually selecting axes that represent the two dimensions of maximum variability. Based on previous work, there seems to be agreement that there will be a high amount of variability in how users access the system (frequency and duration). There also seems to be a high amount of variation in the amount and size of data accessed. Given that a matrix is developed spanning the continuum along these axes, most (if not all) users will fall somewhere in the matrix. Variability in disciplines and type of data used can be reflected within the matrix itself.

2.4 User Scenario Matrix

2.4.1 User Scenario Matrix, December, 1993

After analyzing and discussing the three User Model/Data Model approaches discussed above, the Hughes Team agreed to build upon the work already conducted, while following the concept of a two dimensional matrix that captures the maximum user variability. It was felt that the Ad Hoc Working Group White Paper had a good approach to categorizing users with respect to system access patterns, as did the Hats Paper**.** A new list of system access patterns was developed by merging the 17 categories from the Ad Hoc Paper with the 16 categories from the Hats Paper. Categories whose definitions were very similar were combined; for example, the categories *Algorithm Developer/Maintainer* ("hats" paper) and *Science Software Developer* (Ad Hoc Paper) are defined as (paraphrasing):

Science Software Developer : responsible for integrating software and delivering the production version of the software to the EOSDIS

Algorithm Developer/Maintainer: develops and/or maintains an EOS algorithm for the production of Earth Science data. Includes writing and testing the algorithm code.

This merge process led to a list of 14 system access patterns.

The next issue to be addressed is the data access patterns of the users. The Ad Hoc Paper bases data access patterns primarily on areal coverage of data, while the Barkstrom paper groups users according to type of research activity (case studies, field studies, climatologies, theoretical studies, and reviews). The Barkstrom paper also assigned numbers to each category based on a search of the published literature. After reviewing both documents, the decision was made to use Barkstrom's categories of research activities. These categories coupled with the system access pattern categories form a 14 x 5 User Scenario Matrix resulting in 70 user types. Due to the amount of time required to develop 70 scenarios, the system access categories were re-examined to determine if further consolidation was possible. The system access groups were regrouped based on level of user expertise For more detailed information on the regrouping, see the *ECS User/Data Model Approach and Plan White Paper*, June 1993.

It should be noted that the new system access pattern categories are the same seven final categories as those presented in the Ad Hoc Paper. The resulting User Scenario Matrix (with examples of scenarios) contains 42 elements and can be found in Appendix A of this document. Definitions of each user category are also included in Appendix A. This User Scenario Matrix was in use in the December, 1993 time frame.

2.4.2 User Scenario Matrix, June 1994

In January of 1994, the focus of the User Scenario Matrix changed from the entire EOSDIS user community to Science users only. This decision resulted in fewer user categories; for example, the Science User Scenario Matrix does not distinguish between Graphical Interface users and Character Text Interface users as the prior matrix did. The two principal components of system usage, namely system access patterns and data access patterns carry over to the new matrix. In the process of refocusing the matrix to the science community, the non-science community information and demographics are not lost; their needs were input to the modeling effort via mechanisms other than the User Scenario Matrix. As the matrix itself was refocused, the user classes were redefined. It should be noted that although the user classes are not mutually exclusive in the strictest sense, they can be applied to the *primary* activity that a scientist performs. The June 1994 Science User Scenario Matrix is shown in Table 2-1 and is followed by the user class definitions.

Table 2-1. June, 1994 Science User Scenario Matrix

	Reviews	Local/Field/Case Studies	Regional Studies	Global Studies
Traditional User contacting EOSDIS directly	**Ph D student needs information for dissertation literature review** *David Flittner* *Kathryn Neel* **1**	**Researcher studying lightning associated with flash floods** *Ronald Holle* *Lori Tyahla* **2**	**Test ecological theory regarding vegetation competition in grasslands across the central U.S.** *Don Strebel (Piers Sellers)* *Lori Tyahla* **3**	**International researcher (Scotland) developing Forest Model** *Andrew Friend* *Kathryn Neel* **4**
Data Consumer	**Earth Science Researcher wishes to access electronic journal** *Jeff Dozier* *Lori Tyahla* **5**	**Regional Park Land Management (VA)** *Jerry Garegnani* *Joe Miller* **6**	**Development of method to integrate data sets of varying resolutions.** *Dan Baldwin* *Lori Tyahla* **7**	**Study of Biomass burning** *Chris Justice* *Tess Wingo* **8**
Data Browser	**Undergrad. in Remote sensing class needs info on EOS instruments and Data sets** *Jan Poston* *Tess Wingo* **9**	**Land Surface Hydrologic Model** *Ted Engman* *Joe Miller* **10A**	**Arctic Icepack Response to Weather** *John Heinrichs* *Lori Tyahla* **11A**	**Mid-latitude and tropical interactions - precipitation forcing** *Jim Stobie (Ricky Rood)* *Lori Tyahla* **12**
		Validation of Cloud Properties With Field Data *Bruce Wielicki* *Haldun Direskineli* **(Wingo) 10B**	**Derivation of Snow Water Equivalents** *John Walsh* *Khalsa/Kaminski* **(Wingo) 11B**	
			Radiative Fluxes over sea ice *Jeff Key* *Khalsa/Kaminski* **(Wingo) 11C**	
Analytical User	**Earth Science Community User; e.g., University Prof., Radiation Budget** *Barkstrom (CERES)* *Haldun Direskineli* **(Wingo) 13**	**Development of Automated Snow Mapping Procedure (Sequoia 2000 Scenario)** *Walter Rosenthal* *Lori Tyahla* **14**	**NOAA researcher studying seasonal and diurnal variation in regional lightning distribution** *Raul Lopez* *Lori Tyahla* **15**	**Southern Ocean Large-Scale Circulation** *Leonard Walstad* *Tess Wingo* **16**
Production User		**Watershed modeler Updating model inputs and providing output to EOSDIS** *Jim Hannawald (EPA)* *Joe Miller* **18**	**Biogeochemical fluxes at the Ocean/Atmosphere Interface** *Catherine Goyet (Peter Brewer)* *Kathryn Neel* **19**	**ISI Global Water Cycle; includes model verification through field studies.** *Eric Barron* *Joe Miller* **20**
Machine-to-Machine Interface User		**Thermal Alarm System for Detection of Volcanic Eruptions** *Luke Flynn* *Lori Tyahla* **22A**	**Stratospheric chemistry and dynamics** *Leslie Lait (Mark Schoeberl)* *Lori Tyahla* **23A**	**EOS Instrument Investigator; e.g., MODIS, Ocean Color** *Mark Abbott* *Lori Tyahla* **24**
		Climatic and tectonic processes in the Andes mountains *Bryan Isacks* *Joe Miller* **22B**	**Validation of Passive Microwave Algorithm for Precip. retrieval** *Michael Goodman* *Danny Hardin* **(Tyahla) 23B**	

System Access Pattern Definitions (rows of matrix):

Traditional User: This user interacts on an irregular basis with EOSDIS by speaking with a User Services Person via a telephone. They do NOT EVER access EOSDIS directly. If they phone in once to receive instructions, etc., and plan to log on themselves, they are a different type of user.

Data Consumer: A Data Consumer receives data on a regular basis and DOES NOT return any data to the EOSDIS. A Data Consumer can be a user who phones User Services and sets up a standing order for data. A Data Consumer can also be a ChUI or GUI user who receives data non-interactively on a regular basis. A Data Consumer CANNOT be a user who is bypassing the standard user interface.

Data Browser: A Data Browser is a user who spends much more time locating and choosing his or her data than ordering it. The user does not perform any analysis on the data other than visual inspection.

Analytical User : An Analytical user is a user who does more than locate and receive data while on-line. This user intends to perform data analysis which is beyond visual inspection. Analysis includes subsampling, movie loops, etc.

Production User: A Production User is one who is producing data which is managed by the EOSDIS. The user may be using EOSDIS data as input to their production process, but the purpose is to produce a new data set (or a higher level product) that may or may not be available to all other users. Users in the process of DEVELOPING an algorithm are NOT Production Users.

Machine-to Machine

InterfaceUser: All users who are accessing EOSDIS data via an automated process (no human intervention required) are Machine-to-Machine Interface Users, regardless of what type of activity they are engaged in. For example, if a researcher is using software to search for data or monitor particular data streams, then he or she is a Machine-to-Machine Interface User regardless of the purpose of the data searching or monitoring activity.

Data Access Pattern Definitions (columns of matrix)

Reviews: This category is mainly for document searching and viewing and possibly for obtaining a browse image or small data sample. It is meant for scientists who are doing reviews or summaries of scientific research in particular fields. The author of a review may also look for a "representative" image or sample of data to include in his or her review paper.

Local/Case/Field

Studies: Local and field studies are studies of a region which is geographically small (less than 1000 km^2 or 400 mi^2). A case study is the study of an event which lasts for a season or less, but the geographic extent may be somewhat larger than "local". An example of this is a researcher studying a particular supercell thunderstorm which may be large and also may change position with time.

Regional Studies: A regional study is defined as the study of data that is related to a geographical area which is larger than 1000 km^2 but smaller than 10^7 km^2 (4 x 10^6 mi^2). For example, the Mississippi flood affected a region, not a local area. The volume of data required is typically larger than in a Local/Field/Case study, but smaller than a Global study.

Global Studies: Global studies are defined as studies or processes which are global in scale (the area under study is larger than 10^7 km^2 4 x 10^6 mi^2, studies where the objective is to produce a global data set, and/or studies involving global distribution of parameters. Global studies require the largest volume of data.

A procedure was also developed to assist in placing users into the Science User Scenario Matrix. The procedure consists of a series of questions that are arranged in such a way that if the answer to a particular question is "No", one user class is eliminated and the next question is asked. If the answer to a question is "Yes", the category is then identified. This process is depicted in the form of two flow diagrams , one for the system access classes (Figure 2.1a) and one for the data access classes (Figure 2.1b).

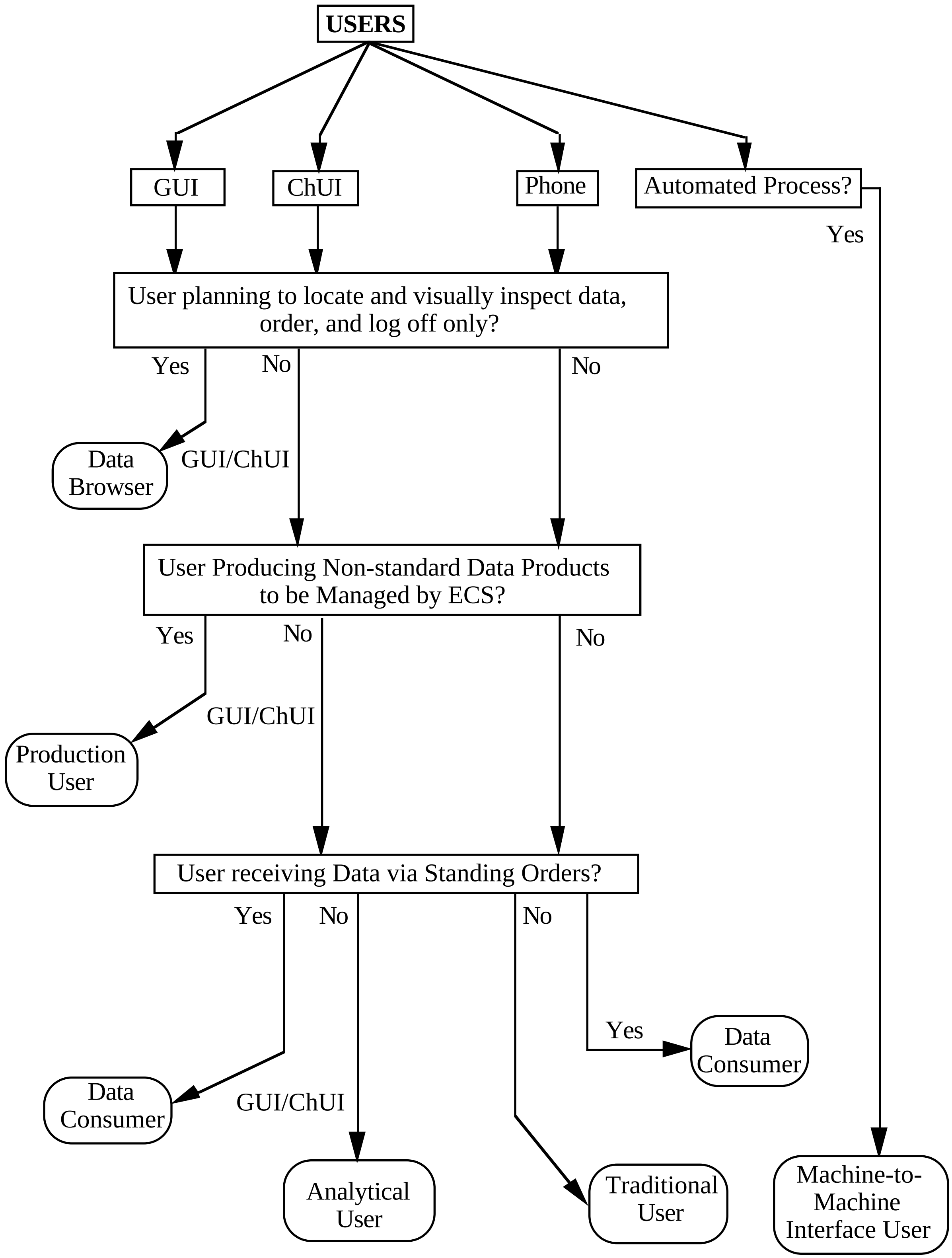


Figure 2.1a. Flow chart to categorize a user's system access pattern

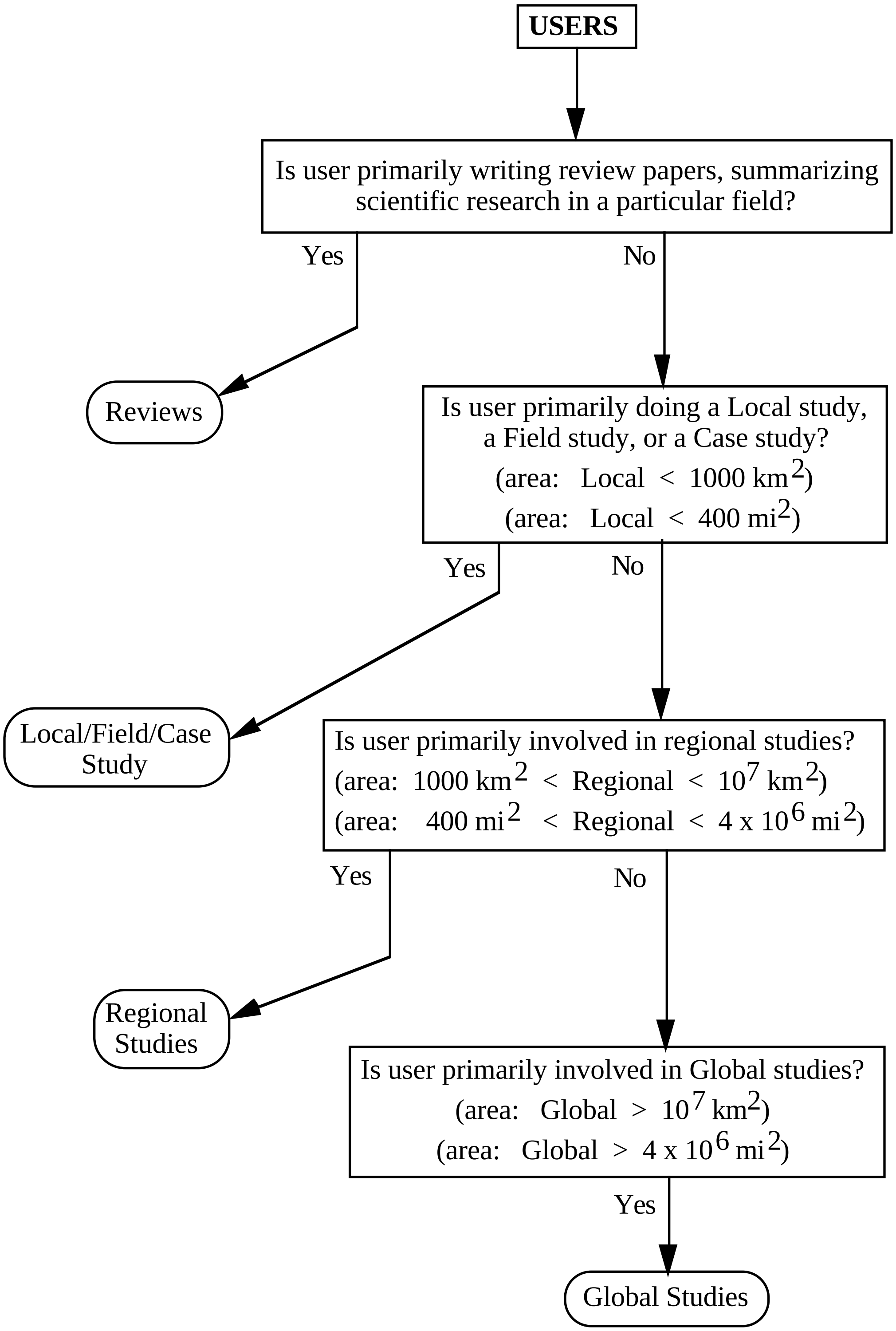


Figure 2.1b. Flow chart to categorize a user's data access pattern

3.0 Scenario Collection and Analysis

3.1 Introduction

The collection of user scenarios and the subsequent analysis of the scenarios is a very important part of the overall ECS design process and serves multiple purposes. One purpose is to ensure that the planned ECS functionality meets the requirements of those who intend to use the EOSDIS. When scenario functionality is combined with demographic information, system and communications network loading can be estimated. In addition, the scenario information is used to build lower-level service threads that depict the order in which services will be invoked by users. Also, the user scenarios provide critical detailed information regarding how users expect to access, browse, and order data. Thus, the collection and analysis of the user scenarios must be a careful process in order to ensure that the information passed to the ECS developers accurately represents the intended interactions of ECS by the user community. This section describes the scenario collection process, the analysis of scenarios and validation of the user matrix.

3.2 Scenario Collection

Scenario subjects were selected from a wide range of investigation types. An even distribution of land, atmospheric, and oceanic science scenarios was sought, at a variety of temporal and spatial scales, so as not to bias the estimated demand for particular types of data . The time frame for each scenario was assumed to be 1998, and the cost of the data was assumed to be a nonfactor. Scientists were encouraged not to limit their desires for functionality because by listening to their desires now, the direction that functional development should take to meet their needs becomes more apparent. Each scenario does not necessarily encompass a scientist's entire investigation, but deals with critical parts of it -- data acquisition and manipulation, data analysis, or a combination of all three areas.

When possible, scenario collection was implemented by visiting the scientist at his or her home facility. In many cases, a team of EOSDIS Core System developers participated in the site visit. Presentations were made on aspects of ECS that were of interest to the scientists and two to three hours of each visit were set aside for a detailed interview with the scientist. From this interview, a stepwise, detailed description of how the scientist might wish to use the EOSDIS to support a particular facet of his or her research was constructed.

Notes made during the interview were typed up in rough draft form with as much stepwise detail as possible. These were sent to the scientist for review and comment. In each case, scenario development occurred in an iterative fashion with the text being reviewed 2 - 3 times by the scenario representative. Revisions to the scenario text which were not possible during the time of the visit were captured via subsequent phone calls and email and facsimile (FAX) communications. As analysis of each scenario proceeded, communication was maintained with the user to ensure accurate interpretation of the user's needs.

3.3 Translation of Scenario Texts to Engineering Data

Once the step-by-step scenario text had achieved a stable, near-final form, the information contained therein was quantified via a *scenario template*. The main purpose of the scenario templates was to aid in the creation of service threads by extracting engineering information from the scenario text. One template was completed for each step of the scenario. In general, each template stated a user's request, the system-level service invoked to fulfill the request, and the results of the request expected by the user. Particular attention was paid to the following attributes:

1. <u>Service</u>: What services will the user want to use to access the type of data or information in which they are interested?

2. <u>Data:</u> What data is accesses in each step and what is the volume of that data?

3. <u>Results of User Request:</u> What does the user expect the results of his or her request to be?

4. <u>Mode of Receipt</u>: How will the data be delivered to the user? Will the user desire electronic transfer, media, or delivery to an account on ECS?

5. <u>Frequency</u>: How many times will the user repeat this step during one "session"? How many times per year is the entire scenario repeated?

3.4 Functional Analysis of User Scenarios

The information collected from users in the form of scenarios was analyzed for new system requirements. If the scenario itself required system features that did not seem to be current design considerations, the requested feature was entered into the Recommended Requirements Database (RRDB). The RRDB administrative team first compares each RRDB record to the existing requirements to determine if the request is a new requirement. In some cases, a requested feature did match existing requirements; in other cases, the request was an implementation detail of an existing requirement. Several of the requested features are new requirements and are sent to the RRDB Screening Team for further action.

In addition to entering system requirements into the RRDB, the system functionality requested by users via the scenarios has been documented in detail in the White Paper, *User Scenario Functional Analysis* , (Doc # 194-00548TPW). This document describes each system function requested by users as well as the current (as of June 1994) disposition of the RRDB record corresponding to it. Also included in the document are references to the System Design Specifications (SDS) document and the Functional and Performance Requirements Specifications document (F&PRS) as they pertain to each requested system function.

3.5 User Scenario Matrix Validation

One of the central goals of the user scenario methodology is to characterize the large and diverse ECS science user community. In order to determine if this goal has been met by the particular choice of scenarios, one must devise a method to verify that these scenarios are representative of the ECS science user community. Recall from Section 2.4 of this document that the two principal components of the scenario matrix are system access pattern and data access pattern. Therefore, these components are used to determine the validity of the matrix.

3.5.1 System Access Pattern

System access pattern can be related to frequency of system access in the following way. A Traditional User (top row of matrix) is expected to access the system infrequently because this type of user is not interested in learning how to use the system; he or she only desires to obtain data. A user in the Data Consumer category is expected to use the system infrequently as well because, by definition, this is a user who determines his or her data set of choice and then receives regular shipments of it without having to re-access the system.

At the higher access frequency end we have Production Users, who will use the system frequently because they are producing a product over time and need input data from ECS. The highest frequency of use will be the Machine-to-Machine Users. These users may have continuous access of the system because the data is accessed by a user's process and does not require human intervention. Thus, to measure frequency of system access, one can use the number of user requests per day. This measure is somewhat complex for Machine-to-Machine users because it will be difficult to quantify a process that is accessing data in a near-continuous fashion.

3.5.2 Data Access Pattern

Data access pattern can be related to volume of data accessed. It is expected that a researcher doing a review paper will access a very small amount of data. This type of user may access several documents for his review paper, but the relative size of documents is small when compared to the size of the data products. A researcher who is performing local, field, or case studies will require a larger amount of data than a reviewer, but the volume will still tend to be small because of the areal coverage of the study.

The data volume required by a researcher performing regional studies may vary greatly for two reasons. If the area of study is small, but high resolution data is required, the user may access a larger volume of data than a researcher studying a larger region but at reduced resolution. In addition, if two researchers are studying identically-sized areas, but one is studying long-term data and the other is interested only in short-term data, the volume accessed by the first researcher will be larger than that accessed by the second. The largest volume accessed by a researcher in the Regional category will occur when he or she is studying a large region with high resolution data over a long period of time.

Researchers in the Global Studies category are expected to receive very large volumes of data. However, some global researchers use data at a very coarse resolution, such as 6^{o} x 6^{o}. Others may be interested in zonally averaged values for the globe which will also result in a relatively small amount of data accessed. Others will want global data at finer resolution resulting in large amounts of data accessed.

3.5.3 User Requests and Data Volumes Delivered to Users

Given the possible combinations of spatial coverage and resolution and temporal coverage and resolution, there will be areas of overlap between the columns (data volumes) of the user scenario matrix. Again, the largest area of overlap will be between the Local/Field/Case Studies and Regional Studies and between the Regional Studies and the Global Studies primarily due to the wide variability in data volume accessed by the Regional Study users. In addition, there will also be small overlap areas between the rows of the matrix as well. It is conceivable that a researcher will call and order data more frequently than a user who accesses the system once to set up a standing order for data. Indeed, this is the case with one of the collected scenarios.

Figure 3.1 shows number of user requests plotted as a function of volume of data delivered to user. The number of user requests is the defined as the number of high level services that the user invokes when he or she accesses the system. For example, if a user searches the guide for information, this is counted as one user request each time he or she does this. If a user browses 40 images, this is counted as 40 user requests because this activity is requesting the browse service 40 times. When a user places an order for data, it is considered to be one request for the order service. The number of requests in each scenario was then averaged over a 365-day year to obtain the number of user requests/day.

In computing data volumes, only data volumes *delivered to the user* are included. This volume includes metadata delivered to the user's screen in response to a search action, text data from documents, browse products, and data products - essentially all layers of the data pyramid that are delivered to the user via any mode. This means that the volume of data pulled from the archive and subsequently subsetted for the user is not included; only the volume of the final, subsetted product is included. Calculating the volume in this way will affect the volume received by the Analytical users and the Machine-to-Machine users the most because data volumes sent to a process that is resident on ECS are not included. The volume of data delivered to the user per day is an average based upon a 365-day year (see Section 5.2 for more details on these calculations).

3.5.4 Matrix Validation Results and Discussion

Before discussing whether the individual scenarios have been placed in the "correct" categories, it is interesting to note that this set of scenarios spans a large range of both data volume delivered to each user and number of requests per day for each user. The upper right corner of the plot does not contain any scenarios; however, this is not unexpected. A user who wants to receive more than 1 GByte of data per day is unlikely to be accessing the system frequently so that his or her number of requests per day will be relatively low.

It is clear from Figure 3.1 that the user scenarios collected thus far span a wide range of user types. Although the data volume and frequency of access for each scenario may not exactly match its position in the user scenario matrix, the scenario matrix is still a valid conceptual tool for classifying the user community.

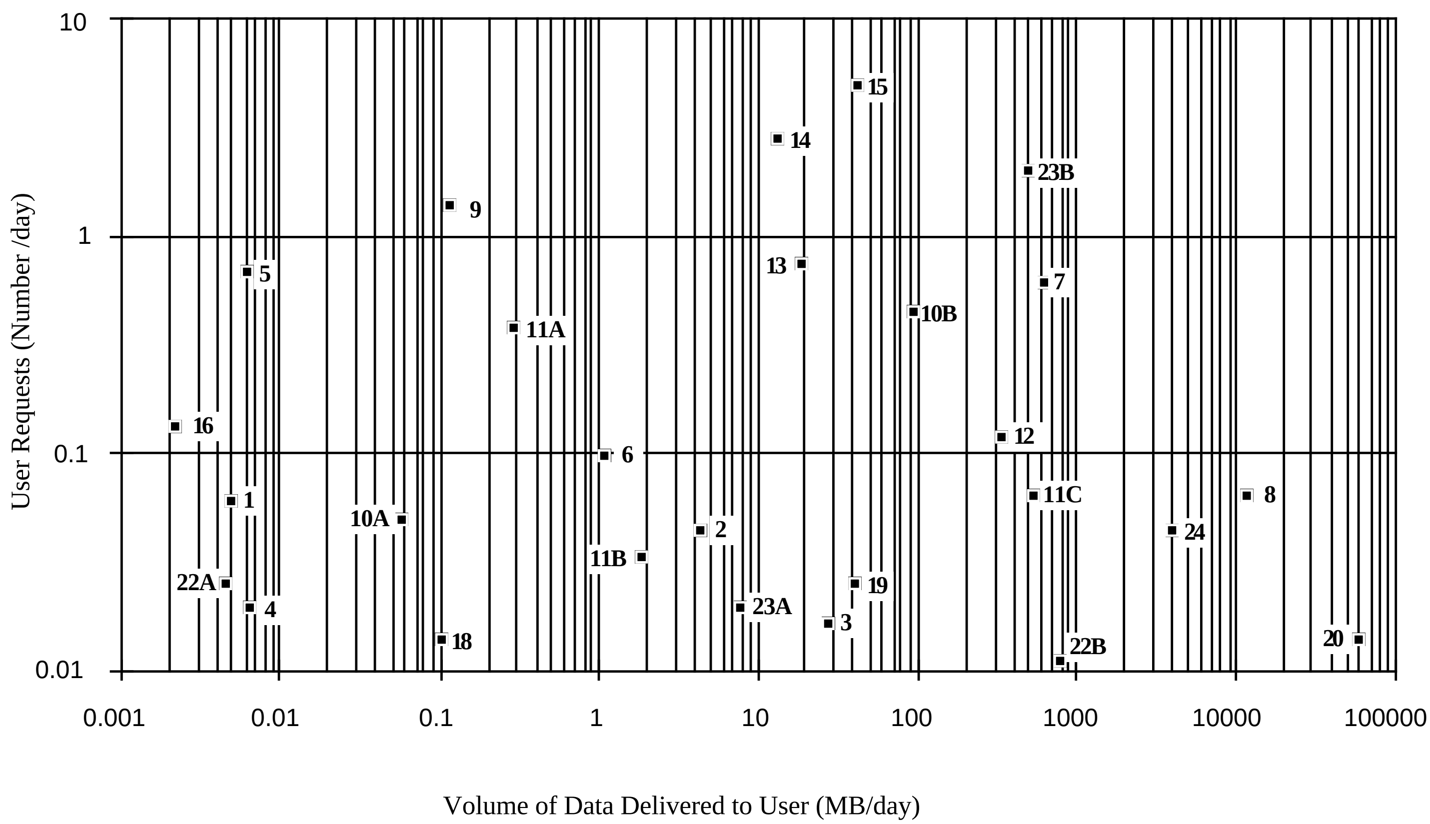


Figure 3.1. User Requests vs. volume delivered to user from user scenarios (both quantities are averages based upon a 365-day year)

3.6 Review of User Scenarios by Science User Working Groups

In addition to the validation activities discussed above, the science user scenarios are in the process of being reviewed by the Science User Working Groups. There is one User Working Group affiliated with each DAAC and the group is made up of members of the scientific community. The purpose of the scenario review by these groups is to ensure that the user scenario collected by the User Characterization Team is fairly representative of the user class to which it is assigned. If a scenario reviewer feels that the collected scenario is not representative of that user class, the User Characterization Team will re-examine the scenario, contact the user for possible alterations to his or her scenario, and decide whether it is both necessary and feasible to pursue an alternate scenario for the user class in question.

To date, the comments regarding the user scenarios have generally been positive and did not require significant alterations of the original scenarios. However, one reviewer expressed major reservations regarding one of the scenarios. The User Characterization Team sent one of its members to meet with the reviewer to discuss possible alterations to the scenario in question and this feedback will be incorporated in the next release of the user scenarios.

3.7 Scenario Inputs to Modeling Activity

The information contained in the scenarios collected from the user community was used as input to other modeling activities, such as the Activation Model, the Data Model, and the Volumetric Model. Close interaction between the scenario collectors and the members of these other modeling groups was required to ensure that the users' needs and research activities were accurately represented. This document provides a description of the interaction of the modeling groups only and does not include detailed descriptions of the models themselves.

3.7.1 Scenario Inputs to Activation Model

The high level purpose of the Activation Model is to act as an interface between the user-oriented scenarios and the engineering-oriented Static Model. At a lower level, there are several goals of the Activation Model. The first is to produce a service thread for each step of a scenario. A service thread describes the lower level system services which must be invoked to fulfill the higher level user request. The second goal is to determine the frequency at which each lower level service will be invoked during the implementation of each scenario step. . A third goal is to identify data products of interest and determine the volume of data operated upon and/or distributed to the user in each scenario step. The activation modeler and the scenario collectors work closely to achieve these goals.

Upon completion of the scenario templates (see Section 3.3) by the scenario collector, a meeting was held with the activation modeler to discuss the scenario text and templates. Each step in the scenario is discussed in detail to maximize information transfer from the User Model group to the Activation Model group. The goal of the scientist's research was discussed to provide the activation modeler with a context to aid in understanding the user's requests and the user's expected results of his or her requests. Specific data products (from the GSFC Science Processing Support Office Standard Data Product List, dated 2/14/94, and non-EOS data) that met the user's needs were discussed and decided upon. Knowledge of the instrument launch schedules was also required in order to ensure that a user is not accessing data before it is available from the instrument.

Based upon the discussions with the scenario collectors, the activation modeler produces a "supporting analysis" for each scenario. This analysis contains the detailed calculations that result in the data volumes (per day, averaged over a 365-day year) accessed and distributed in each scenario step as well as the number of times that a high level user service is invoked (also per day, based on a 365-day year). The volume of data accessed and/or distributed is calculated based on a hierarchy of data subsetting types. The first volume calculated assumes that entire data granules are pulled from the archive. A new volume is then calculated by subsetting the data granule by parameter of interest. Then this data volume is further reduced by spatial and temporal subsetting to match the requirements of the user. The final volume calculated thus represents the fully subsetted data volume delivered to the user. The analysis also contains calculations which produce the number of requests per day (averaged over a 365-day year) for each scenario step.

The activation modeler then reviews his analyses with the scenario collectors to ensure that the analyses accurately represented the users' expected interactions with EOSDIS. Occasionally, additional information is required from the user to in order to complete the analysis; in this case, the scenario collector contacts the user for clarification and then relays the information to the activation modeler. The activation modeler then builds service threads for each scenario from the analysis in a format that is compatible with that required by the Static Model.

3.7.2 Scenario Inputs to Volumetric Model

The objectives of the Volumetric Model are to describe the distribution of data server loading that results from the delivery of Browse and Level 0 through Level 4 data products on a DAAC-by-DAAC basis to all users and to validate the results of the Static Model for the General Science user category. It is not the intent of this document to describe the Volumetric Model in detail; the intent is to document the methodology employed by the User Characterization Team to produce the inputs to the Volumetric Model.

The User Characterization Team provided the developers of the Volumetric Model with the detailed data requirements from the user scenarios. Each step in a user scenario that either accessed Level 0 through Level 4 data and/or Browse data was provided with the data product name, spatial extent and resolution, and temporal extent and resolution. The number of users that each scenario represents was also provided (see section 4.3).

3.7.3 Scenario Inputs to Data Model

The two main high-level purposes of the Data Model are 1) to provide one consistent reference for information that is deemed design-critical, so that all parties can understand the ECS data and information at a certain level of detail and 2) to generate the highest (conceptual) and intermediate (logical) level models of data that can be converted into low-level physical implementation plans by those who will perform detailed design and, ultimately, implementation of the various elements of the ECS software. This activity is very important as the results may be used to maximize the efficiency of both data access by users and the processes by which standard products are produced. The current (as of June, 1994) Data Model is based upon inputs from many groups and activities; one key input was the science scenarios. Upon completion of the scenario texts and engineering templates, the scenario writers met with the data modelers on a one-to-one basis to discuss the scenario information in detail.

At each meeting between the scenario writers and the data modelers, one scenario was discussed step-by-step. For each scenario step, the following information was recorded: request type, request criteria class, request criteria, service invoked, data retrieved, and result type. This information was then used by the data modelers to define the objects contained in each data pyramid layer and attributes of these objects and to develop logical object data elements and logical object classes.

4.0 EOSDIS User Demographics

4.1 Introduction

Knowledge of the potential size of the ECS user community is critical to several aspects of the system design, including network and data server loads. The methodology for arriving at estimates of the size of the user community relies on several sources of information and the methodology itself has evolved since the inception of this activity. This section describes the methodologies used to collect demographic information prior to the System Review in December of 1993 as well as those used to support the System Design Review (SDR) in June, 1994. Pre-December Review methodologies are described first because some of these activities continued in support of the SDR in June, 1994.

At a very high level, the ECS user community is divided into three main categories: EOS Science users, General Science users and Non-Science users. EOS Science Users are the investigators funded by NASA under the EOS program. The category of General Science users includes all science users. This category includes, but is not limited to, EOS-funded researchers, university researchers, and federal employees who conduct basic Earth System research. The category of Non-Science users includes the widest variety of users, ranging from commercial users to K-12 users.

Section 4.2 describes demographics of the EOS Science user community, section 4.3 describes the General science community demographics, and section 4.4 describes the demographics of the Non-Science user community. Each major section details the methods used to arrive at the estimated number of users and the results of those methods. Section 4.5 contains a summary of the results in tabular form.

4.2 EOS Science User Community

The demographics of the EOS Science user community are well-defined because this group includes only those investigators who are funded by NASA under the EOS Project. Included are all Principal Investigators as well as the Team members of all of the Interdisciplinary Science Investigations and all Instrument Investigators and Team members. The current (as of 20 April, 1993) total number of these investigators is 633. It is further assumed that each of these investigators will have, on average, 3 additional staff working on these investigations.

The EOS science community from other countries was derived in the same way as for the EOS Science category in the United States. The estimates for the General Science investigators from other countries are very preliminary and were based upon a survey conducted in Europe for European users of Earth observation data; estimates for other areas of the world are no more than subjective estimates by the contributors to this White Paper. If the size of this community is deemed to be important, more work is required to arrive at a better estimate.

4.3 Demographics of the General Science User Community

Original work for estimating the size of this portion of the earth science community was done by Joseph Miller in support of the System Review in December, 1993. The emphasis in this approach was to determine demographic estimates for each cell of the User Scenario Matrix (December version, see Appendix A). Two techniques were employed based upon the work of Dr. Bruce Barkstrom (*A Preliminary EOSDIS User Model*, July 1991) - estimation of demographics from examining 1) professional society membership and 2) professional journal publications. In addition, an independent method based upon information in *Peterson's Guide to Graduate Programs in the Physical Sciences and Mathematics* (1994) was employed to validate the previous results.

4.3.1 Demographic estimates from Professional Society Membership (Miller December 1993)

The following professional societies were contacted for membership information: American Geophysical Union (AGU), IEEE Geoscience & Remote Sensing, American Society of Agronomy, American Meteorological Society, and Geological Society of America. Membership data were provided for 1992. In addition, five professional journals were studied: *Journal of Geophysical Research (JGR) - Atmospheres, JGR - Oceans, AGU - Water Resources Research, IEEE - Geoscience and Remote Sensing*, and the *International Journal of Remote Sensing*. Since refereed journal publication is a common employment criterion for scientists, examining journal articles is a reasonable input in determining the number of active scientists.

For each professional society an estimate of the actively publishing proportion of members was derived by counting the number of authors in one year of each society's main publication and dividing this number by the total membership in that society. These estimates for active members in each society were added together for an estimate of the number of people active in earth science (see Table 4-1).

It is recognized that there will be some overlap of society memberships as well as in journal publications but the overlap should not greatly influence the total number of active researchers. Based upon this analysis, the estimated number of General Science users is 3,500-10,000.

4.3.2 Demographic Estimates From Examination of Professional Journal Publications

In order to obtain demographic estimates for each cell in the User Matrix, it was necessary to categorize the investigations being conducted by the professional society members. This task was accomplished by a rapid survey of articles published in one year (1990) of the five professional journals listed in the previous section.

A data collection sheet was developed to record pertinent information about each journal article (Table 4-2). This sheet was structured around a truth table of viable combinations of spatial and temporal coverages and resolutions of remotely-sensed data. Different combinations of these characteristics were deemed unlikely or impossible and were removed from the table. Other combinations were associated with particular scales of investigation (Reviews, Local/Field/Case Studies, Regional Studies, Global Studies) as identified on the User Scenario Matrix.

Table 4-1. Estimated General Science User Population From Examination of Professional Society Membership

Professional Society	Total Membership	Proportion of Active Members	Number of Active Members
American Meteorological Society	10,300	0.18	1,854
Ecological Society of America	6,551	0.07	459
American Society of Agronomy	12,600	0.04	516
Geological Society of America	16,801	0.025	420
IEEE Geoscience & Remote Sensing	2,560	0.08	205
American Geophysical Union	30,600	0.09	2,754
TOTAL	79,412		**6,208**

When an article was examined each data set referenced in the article was categorized using the truth table. The use of in situ data was also recorded. The various categories of user interface, as developed for the User Matrix, were incorporated into the data collection sheet. Estimates of the type of interface the authors of an article would be likely to use were made by noting the type of investigation, level of technology, and the author's place of employment. If an article was of the "Review" type, this was noted on the categorization sheet for that article. Finally, the number of authors was also recorded. Each author was weighted equally in the final count because each author was assumed to be equally capable of being a principal author of a journal article.

Every other article in each journal issue was examined due to time constraints. Counts from this sample were then doubled to get an estimate for the entire publishing population for those journals. These counts were sorted and categorized using the User Model scenario matrix. What resulted was a count of authors within each of the scenario matrix cells, from which proportions for each journal examined were derived for the year 1990. Matrix cell counts for all the journals examined were added together to arrive at a total count of active researchers which were categorized according to the matrix. From this, estimates of proportions of researchers from the entire earth science community per matrix cell were derived (Table 4-3). Style of system usage is described along the y axis (rows), while scale of research is described along the x axis (columns).

Table 4-2. Literature survey data collection sheet

TYPE OF RESEARCH	DATA ACCESS PATTERN					Article No.	Journal
	Spatial Coverage	Spatial Resolution	Temporal Coverage	Temporal Resolution	# of Sensors or Products		
Climatology	H	H	H	H	H, L		
Climatology	H	H	H	L	H, L		
Climatology	H	L	H	H	H, L		
Climatology	H	L	H	L	H		
Climatology	L	H	H	H	H, L		
Climatology	L	H	H	L	H, L		
Climatology	L	L	H	H	H, L		
Climatology	L	L	H	L	H		
Field Study	H	H	L	H	H		
Field Study	H	H	L	L	H		
Field Study	L	H	L	H	H		
Field Study	L	H	L	L	H		
Case Study	H	H	L	H	L		
Case Study	H	H	L	L	L		
Case Study	L	H	L	H	L		
Case Study	L	H	L	L	L		
Theoretical	H	L	L	H	H, L		
Theoretical	H	L	L	L	H, L		
Theoretical	H	L	H	L	L		
Theoretical	L	L	H	L	L		
Theoretical	L	L	L	H	H, L		
Theoretical	L	L	L	L	H, L		
					Insitu data		

H: Coverage greater than regional	H: Detail greater than 2 deg x 2 deg	H: Coverage longer than seasonal	H: More frequent than weekly	H: More than two

No. of Authors	

Character Text	
Data Consumer	
Data Browser	
Analytical User	
Production	
Advanced Tech.	

Table 4-3. User Model matrix proportions for Earth Science community

	General Info.	Reviews	Theoretical	Case Study	Field Study	Climatology/ Global
Character Text	0.00	0.00	0.01	0.03	0.02	0.00
Data Consumer	0.00	0.02	0.00	0.03	0.01	0.02
Data Browser	0.00	0.04	0.00	0.01	0.02	0.01
Analytical User	0.00	0.03	0.12	0.23	0.15	0.13
Production	0.00	0.00	0.00	0.03	0.04	0.04
Advanced Tech.	0.00	0.00	0.00	0.00	0.00	0.00

These earth science community matrix cell proportions were then applied to the estimated total active membership estimate of 6,208 (Table 4-4).

Table 4-4. User Model matrix estimates for the Earth Science community

	General Info.	Reviews	Theoretical	Case Study	Field Study	Climatology / Global
Character Text	0	14.72	44.17	166.86	122.69	29.45
Data Consumer	0	98.15	29.45	166.86	58.89	112.87
Data Browser	0	240.7	9.82	73.61	152.13	58.89
Analytical User	0	191.39	741.03	1428.09	937.33	819.55
Production	19.63	9.82	19.63	176.67	260.10	225.75
Advanced Tech.	0	0	0	0	0	0

Note that the previous two tables include only those categories of system access style which were thought to be discernible by examining published research articles. The categories of system access style or "user style" of 'Intermediary' and 'Traditional User' are not included because they describe a commercial vendor (unlikely to appear in a research journal), and a scientist who would telephone for data rather than log onto the system (impossible to discern by looking at a journal article).

4.3.3 General Science User Demographic Estimation Update (Stanley - April 1994)

The demographic estimates made in December 1993 were updated in April 1994 by Thomas Stanley using methods similar to the Miller study. More current membership information (for the year 1993) was obtained from the *Encyclopedia of Associations*.. Author counts from journal indices were made for all available journals to determine the proportion of society membership that was actively publishing. To account for authors who were missed because they were in other journals of the same society that we were not able to obtain, the count from the available journal was multiplied by the total number of journals for that society. For example, only one of the research journals published by the American Meteorological Society was available for examination, therefore the count for *Journal of Climate* was doubled to account for the researchers publishing in *Physical Oceanography*.

The proportions of actively publishing researchers were calculated by dividing the number of authors counted for a particular society by the total number of active authors for all earth science-related professional societies. This was essentially the same method that was used in December 1993, but was applied to the actively publishing segment of the membership. To estimate the number of potential EOSDIS users , Barkstrom's (1991) method of applying 5/8, 1/2, and 3/4 to estimate the lower bound, the mean , and the upper bound of the range respectively, was employed (see Table 4-5).

4.3.4 Adaptation of demographics data to modified User Matrix (March 1994)

The User Model scenario matrix was re-examined after the December Review. At this time, it was decided that the User Matrix would contain science user scenarios only; other types of users would be included via other methods. This decision resulted in fewer user classes; thus, the December 1993 demographic information for the science community was recalculated.

The journal data collection sheet was also re-examined and the following categories of data access were re-mapped to the new User Model matrix (Table 4-6). Note that the "Reviews" category was not re-mapped because the types of data used by scientists writing review-type articles are mainly tabular and bibliographic in nature.

The matrix cell proportions developed in December 1993 were remapped to the new User Matrix by re-examining all of the original journal data collection sheets. These were sorted according to Table 4-6, and new proportions were calculated for the new matrix (Table 4-7).

Table 4-5. Comparison of demographic estimates

Process Step	Barkstrom	Miller	Stanley	Notes
Society Membership	72,200	79,412	76,160	
Journal Authorship		6,208	7,337	• Journal authorship is an author count from the journals • Total authorship in society is the number of journals in Society multiplied by the number of authors determined in the author count. • Total Publishing Authors is the sum of each society total authorship
Range of # Publishing in Society				
Lower Bound	0.08	0.03	0.04	Range of publishers is based on fraction of society
Mean	0.1	0.1	0.1	membership that publishes
Upper Bound	0.18	0.18	0.24	
Population				Based on above range of proportions.
Low	3,800	2,382	3,046	
Expected	7,200	7,941	7,616	(Range limit)(total society membership)
High	13,000	14,294	18,278	
Estimated # of Publishing Researchers Interested in EOS data				Estimate of Potential EOS Customers based on Barkstrom's estimated fractions (1/2,5/8, and 3/4)
Lower (1/2) Low	3,000 (pub.) *2,900 (act.)*	1,191	1,523	Barkstrom's published number is a rounded number
Expected (5/8) Expected	4,500	4,963	4,753	
Upper (3/4) High	10,000 (pub) *9,750 (act.)*	10,720	13,708	Barkstrom's published number is a rounded number

4.3.5 Independent Validation Technique

In order to validate the previous results for the General Science user category, an independent technique was developed based upon information in the 1994 issue of *Peterson's Guide to Graduate Programs in the Physical Sciences and Mathematics* . This handbook categorizes departments at colleges and universities according to Atmosphere, Ocean, and Earth Science, among others. The handbook contains detailed faculty descriptions for some facilities, but not all; some departments reported the total number of faculty with no details regarding areas of research.

Table 4-6. Mapping of journal data sheet categories to new matrix categories

Journal Data Collection Sheet Categories	New User Model Matrix Data Access Categories
Climatology-type Studies with 'High' spatial coverage	Global
Field Studies with 'High' spatial coverage	
Case Studies with 'High' spatial coverage	
Theoretical Studies with 'High' spatial coverage	
Climatology-type Studies with 'Low' spatial coverage	Regional
Theoretical Studies with 'Low' spatial coverage.	
Field Studies with 'Low' spatial coverage	Case/Local/Field Studies
Case Studies with 'Low' spatial coverage	

Table 4-7. New user model matrix proportions for the Earth Science community

	Reviews	Local/Field/Case	Regional	Global
Traditional	0.16 %	0.81 %	0.16 %	0.25 %
Data Consumer	1.21 %	10.0 %	5.00 %	7.00 %
Data Browser	3.31 %	15.0 %	5.00 %	7.00 %
Analytical User	2.74 %	7.00 %	4.00 %	10.75 %
Production	0.16 %	5.81 %	1.69 %	5.00 %
Machine-to-Machine User	0.81 %	0.24 %	2.90 %	4.00 %

First, the departments listing detailed faculty information in each of these three disciplines were examined. The research interests of each faculty member in these departments were used to determine if they might make use of remotely-sensed data. Each faculty member likely to use remotely-sensed data was counted and recorded. In addition, the total number of faculty in each of these departments was recorded. Then, a sum was taken over the count for each department for the number of faculty using remotely-sensed data to arrive at a total number of faculty using remotely-sensed data in each of the three disciplines. Also, the total number of faculty in each department was summed to obtain the total number of faculty in each discipline. The ratio of these two quantities was then taken to obtain the overall fraction of the number of faculty who potentially will use remotely-sensed data in each discipline.

For each discipline, the proportion of faculty members likely to use remotely-sensed data was applied to all of the departments listed in the Guide in the three disciplines of interest, including those that gave no detailed information regarding the faculty. The result is the number of faculty members that will use remote sensing data in the three disciplines of Atmosphere, Ocean, and Earth Science. It was then assumed that each of these faculty members will have between 2 and 4 graduate students will are also likely to access EOSDIS in their work. The result of this analysis is 2500 General Science users for these three disciplines only.

Since these three categories in *Peterson's Guide* do not include all of the potential science users in Academia, the active membership (5.4% or 1200 people) in three professional societies (IEEE Geoscience and Remote Sensing, Ecological Society of America, and the American Society of Agronomy), which may not be captured in the above estimates ,were added to the above numbers to arrive at a lower bound of 4200 users. This number was doubled to obtain the upper bound of 8400.

The Miller and Stanley estimates included EOS investigators; thus it is necessary to add to the estimate from *Peterson's Guide* the number of EOS investigators. This results in 6,100 to 11,600 potential General Science users of EOSDIS. These results are within the range of values derived via other methods (see Table 4-4).

4.4 Non-Science User Community

The Non-science user category contains the most variability in users. It includes users from Federal and State agencies who are not doing scientific research as well as K-12 students and teachers, commercial users, intermediaries, and library users. Commercial users are different from Intermediary users in that they use the data in-house, whereas the primary function of an Intermediary user is to customize and repackage the data for distribution to end users. This user category has the potential to be much larger than the previous two categories and its impact on system performance must be assessed.

Because of the diversity of this category of users, it was formally subdivided into the following sub-groups: Federal, state, and local government users; commercial users, educational users, and library users. The members of each user subcategory and the methods used to obtain demographic information are described in sections 4.4.1 through 4.4.5.

It is important to note that it was assumed that the data products available to the non-science community are the same data products that were designed for the needs of the scientific community. In some cases (for example, federal government users) surveys were used to determine the data needs of users; in other cases, judgements were made about the degree to which each of the product levels would be of direct interest to typical non-science users. As the time horizon is extended beyond this period, it is expected that the size of the Federal, State and Commercial user community will grow significantly as application techniques and models are developed to enable these communities to apply the EOS data products to their specific needs. The judgements made and the rationale for each are included in the following subsections arranged by user subcategory.

4.4.1. Federal and State Government users

This subcategory is made up of federal and state employees that are not performing scientific research. Based on previous experience, federal, state, and local government users will favor Level 1B image data because they generally apply different algorithms than those of interest to the scientists. For non-image data, the demand will be greater for Level 2 and Level 3 products.

The demographic information obtained for the Federal, State, and Local Government users was based upon questionnaires and interviews. The questionnaire (see Appendix B) was developed for the federal users, but was also employed for state users in some cases. In addition, interviews and workshops were held with both federal and state users.

4.4.1.1 User Questionnaire

A questionnaire was developed to obtain a better understanding of the future needs of the Federal Government agencies for data that will become available through EOSDIS in terms of numbers of individuals and the frequency with which they will want to:

- access the system to review the catalogue of data products available (i.e. to "query" the system);
- review (on-line) samples of the available data products (i.e. to "browse");
- order products and the likely volume of data to be ordered.

The questionnaire was fairly detailed and required a considerable understanding of satellite and Earth Science data by the individual completing it. The questionnaire was first tested with the Department of Interior and feedback was incorporated into subsequent versions. After testing the survey it was decided that individuals within federal agencies who possessed the required data knowledge and were in a position to evaluate their organization (or related segments of their organization) were best suited to complete the survey. The questionnaire was distributed to the Department of Interior, the Department of Energy, the National Oceanic and Atmospheric Administration (NOAA), the United States Department of Agriculture, NOAA data centers, and the existing DAACs.

Although certain biases arise as a result of one individual completing a survey for many people, this method was a good first approximation of the size of this community. In addition, if further estimates are solicited from this community, these survey results may be used for comparison. During the analysis of the questionnaires several follow up phone calls were made to verify and/or clarify information that was returned by the agencies.

Although the questionnaire was originally designed for federal government users, it was tested on several state agencies in Ohio, Alaska, and Texas where adequate understanding of remotely-sensed data was evident. Ohio and Alaska found the questionnaire to be adequate and a good response was received from the agencies in these states. Texas, on the other hand, did not find the survey appropriate for their needs. As a result, a workshop in Houston was set up with many state agencies to discuss their potential interactions with EOSDIS.

4.4.1.2 User Interviews

Several interviews will the Non-Science community to understand issues that may affect the use of EOS data by this community. Several state agencies have found that contracting certain aspects of their work out to local universities is more cost-effective than performing the work themselves. A meeting was held with a researcher at Towson State University to understand how this contracting activity might impact the use of EOSDIS. An interview with EOSAT was also conducted to discuss their interaction with the states in the Statewide Purchase Program that enables state agencies to pool their funds to buy Landsat data.

Considering the above sample information, an analysis was done for all other states in 1992 using information found in the *State Geographic Information Activities Compendium* by Lisa Warnecke (product of the Council of State Governments) to determine the extent of the use of remote sensing and GIS data. From that analysis, the number of potential users per state using data and information similar to that produced by the EOS instruments was extrapolated.

4.4.1.3 Results of Analysis

The data for the federal and state users resulting from the methods above was analyzed in conjunction with prior experience with these communities. The results are tabulated in Table 4-8.

Table 4-8. Potential Number of Federal and State Agency Users of EOSDIS

Federal Agency	Number of Users (Min)	Number of Users (Max)	State Agency Size Category	Number of States	Number of Users (Min)	Number of Users (Max)
DoI	300	400	Large (50-100 users)	17	850	1700
NOAA	600	700	Medium (30-60 users)	15	450	900
DoE	300	400	Small (10-20 users)	18	180	360
USDA	100	200				
EPA	20	100				
DoD	150	350				
Other	30	50				
Total	**1500**	**2200**	**Total**		**1500** (1480)	**3000** (2960)

4.4.2 Commercial Users

Commercial users include companies that have the capability to make direct use of EOSDIS science data products to support business and operations and various planning activities; they use the data in-house. Also included in this category are commercial intermediaries serving organizations that do not have in-house resources for data processing and analysis. Since EOSDIS data products are designed for use in scientific research, they rarely will find direct and immediate application in the commercial world without additional processing and analysis of the data. Consequently the "end user" category includes only those companies that have the interest and resources to tailor products to their needs. This includes companies such as utilities, energy exploration and production companies, agribusiness, and major manufacturers and processors.

Based upon experience with previous remotely-sensed research products, this community will be small, particularly in the early phases of EOSDIS, until the necessary R&D has been completed for techniques to apply the data to non-science applications. The estimates of 100-200 commercial end users and 250-350 commercial intermediaries for the 1998 time frame are based upon discussions with individuals at EROS Data Center (EDC), The Space Remote Sensing Center at the Stennis Space Center, and the Center for Mapping at The Ohio State University, all of whom have had direct experience in working with these communities.

The numbers of Education intermediaries were estimated with the assistance of data contained in the document: “Media Producers of CD-ROM/ Videodiscs” from the National Science Teachers Association of Science Education Suppliers, (1993). Their marketing information helped to determine whether the company was considered to be “large” or “small”. The estimated number of users in each of the categories is as shown in Table 4-9.

Table 4-9. Potential number of Commercial EOSDIS users

Commercial Size Category	Number of Companies	Number of Users (Min)	Number of Users (Max)
Large (3 - 5 users)	15	45	75
Small (1 - 2 users)	32	32	64
Total		**80** (77)	**140** (139)

4.4.3 Educational Users

The current Administration (e.g., the Vice President and the NASA Administrator) have stated strong interest in expanding the use of on-line services and database availability to the education community with programs and funds, these projections might be very much on the low side. Members of Congress and the Education Secretary are proposing "---a federal policy that would ensure that schools are not bypassed as cable and telephone lines are installed for the electronic highway". It is also interesting to note that the use of Internet as of August 1993 by the K-12 community included a total of 111,000 addresses, as estimated by Tony Rutkowski when he was with Sprint International. In this analysis, the Educational user community includes K-12 students and teachers only; the undergraduate and graduate student population who are not members of a scientist's staff are not included. Additional research is necessary in order to quantify the remaining undergraduate user population.

The number of K-12 teachers in specialized science and social science fields such as Earth Science, Environmental Science, General Science, Physical Science, Physics, and Geography that would be interested in EOSDIS products, is estimated to be 53,000 in 1992 (*Market Data Retrieval Educational Mailing List and Marketing Guide*, Market Data Retrieval, 1992-1993). In 1998, the number is estimated to be 56,000 based on a 6% growth in community size (*Projections of Educational Statistics,*. U.S. National Center for Education Statistics, biennial). The number of teachers expected to use microcomputers and networks in 1998 is 50,400 (*Current Population Survey, October 1984 and 1989* , U.S. Department of Commerce, Bureau of Census, Unpublished data). For the purposes of this study, the number of teachers expected to use microcomputers and networks will be used since teachers who will use EOSDIS will probably come from this group. In addition, we estimate that only 5-15% of these 50,400 teachers will use new technologies. Thus, in 1998, we estimate that 2,520 - 7,560 K-12 teachers will make direct use of EOSDIS.

To estimate the size of the K-12 student user community, it is expected that there will be 23 students per teacher on average (*National Science Teachers Association Report*, *1993*) resulting in 58,000-174,000 potential EOSDIS student users. It is interesting to note that this estimate for K-12 student users is several orders of magnitude smaller than the estimated size of the total K-12 student population in 1991 of 46,688,272 (*Market Data Retrieval Educational Mailing List and Marketing Guide*, Market Data Retrieval, 1992-1993). Also, the Market Data Retrieval projections, which are based on data for 1991, do not include any increase in student enrollment due to the construction of new schools.

In summary, the number of EOSDIS users in the Education community compared with the estimated 1998 populations of all K-12 teachers and students are tabulated in Table 4-10.

Table 4-10. Potential Number of Educational users of EOSDIS

Conservative Estimate	**Number of Users (Min)**	**Number of Users (Max)**	**Total Population in 1998**	**Number of people**
K-12 Teachers	2,520	7,560	K-12 Teachers	56,000
K-12 Students	58,000	174,100	K-12 Students	46,688,272
TOTAL	**59,520**	**181,660**	**TOTAL**	46,744,272

4.4.4 Library Users

In a recent editorial in the Washington Post, August 1, 1994, Hardy R. Franklin, past president of the American Library Association, put forth the argument for connecting every public library to the "information superhighway" to give all citizens access to valuable databases.

The estimated number of libraries that will have the capability and interest in accessing EOSDIS is based upon statistics contained in *Library Mailing Lists 1992-1993*, (Market Data Retrieval 1992-93). There are 9,454 Main Public Libraries, and 6,648 Branch Public Libraries for a total of 16,202. However, it is assumed that only those with a book budget of over $20,000.00 will be able to provide access to EOSDIS for the communities they serve. This assumption results in 6,200 main and 3,400 branch libraries. When the estimated number of college and university libraries are included, the range of users associated with libraries that may make use of EOSDIS is between 6,000 and 12,000.

4.4.5 Service Providers

The estimates of the non-science community just discussed assume direct access to EOSDIS. However, another alternative for non-science usage of EOSDIS exists. As the demand from the non-science communities increases, other organizations may step forward to assist NASA in serving certain markets. For example, the Department of Interior's EDC, some of the NOAA data centers or commercial organizations may opt to assist in serving the needs of federal organizations, state agencies, commercial end-users, and intermediaries. NSF, NASA, the Office of Education or commercial enterprises may choose to establish a service tailored to the needs of the K-12 community. Similarly, commercial, county, or state organizations may decide to support the needs of librarians and the communities they serve.

Therefore there may be a select number of service/value-added data providers that interact with EOSDIS to provide products tailored to the needs of specific non-science communities. In the 1999-2003 time frame, the number of potential service/value-added providers (organizations) and the communities they may serve are depicted in Figure 4.1. Generally the flow of data products will be from EOSDIS to these servers; however, federal and state organizations may produce data products of value to NASA's science investigators. Therefore there will probably also be a flow of data from federal and state agencies into EOSDIS.

The nature of the demand on EOSDIS from the value-added providers will be strikingly different than if the individuals from these communities entered EOSDIS directly. The value-added providers will tend to make greater use of machine-to-machine interfaces with EOSDIS and conduct business on a standing order basis, tailoring the EOSDIS products to the needs of various non-science communities and making use of non-EOS data as necessary to meet customer needs.

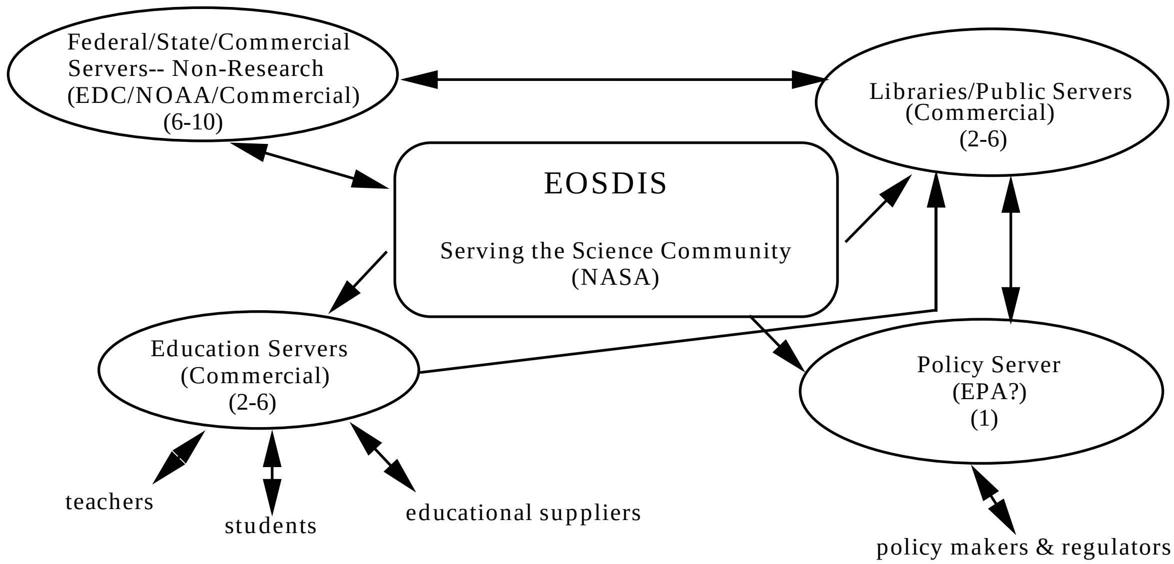


Figure 4.1. Relationship between service providers, users, and EOSDIS

The number of routine EOSDIS users from these service providers can be reduced to 100-300 users rather than the expected 70,000-200,000 from the community at large. These numbers were derived by projecting the number of service providers in the 1999-2003 time frame and the number of user service personnel associated with each provider that would be accessing EOSDIS to obtain products needed to serve their constituents. These estimates are shown in Table 4-11.

Table 4-11. Number of Non-Science Users Serving as Intermediaries to the Non-Science Community

Non-Science User Subcategory	Number of Providers	User Service Personnel for each Provider	Number of users
Federal/State/Commercial	6-10	10-15	60-150
Education	2-6	10-15	20-90
Library/Public	2-6	10-15	20-90
Total Number of Users			**80-330**

The number of users will depend on the nature of services provided by the value-added organizations. For instance, if commercial service providers were to supply only an enhanced interface to EOSDIS and no value-added products, the size of the community accessing EOSDIS directly would most probably be greater than 70,000-200,000. However in this analysis, we will assume a lower number of users based on the assumption that these providers do not simply develop an interface to EOSDIS, but provide data products tailored to meet the needs of the individual non-science communities.

4.5 Summary of ECS User Community Demographics

The potential size of the ECS user community is the sum of the number of users in each user subcategory previously described. Table 4-12 summarizes the results of the demographic analyses for the case where all users are accessing EOSDIS directly.

Table 4-12. Summary of Demographics of EOSDIS User Community

User Community	Minimum Number of Direct Users	Maximum Number of Direct Users
EOS Science Users	1,900	3,200
General Science	4,200	8,200
Non-Science	69,430	268,320
TOTAL user community	**75,530**	**279,720**

5.0 System Access Characteristics

5.1 Introduction

In the context of this paper the term ‘access’ means a system level entry by a user connecting his or her client software to the Data Server or Advertising Service (for further information see SDS Section 4.5.2.3.1) as appropriate. This section first defines the access parameters which have been characterized; then presents the methodologies and results for the Science user community and the Non-Science user community.

5.1.1 Access Frequency

The frequency with which users are expected to access EOSDIS will be a factor in the overall performance of the system. As the frequency of system accesses increases, the speed with which a response to a user request can be fulfilled will decrease if all other system variables remain constant. Thus, it is important to characterize the frequency with which user requests are received by the system.

5.1.2 Access Methods

The methods that will be used in accessing ECS will not vary to any great extent with the user community. The trend is definitely towards the use of on-line or electronic access and, where routine access to large quantities of data is desired, direct machine-to-machine transfer will be employed. Also, since some components of the user community will be associated with other data and information systems, they would access ECS through these systems.

Access methods are important because they define what the various load accesses are and the loads on the system. They also help define what services are needed to support various modes of access, and provide insight into user environments. The fact that other data centers and individuals from other countries will be accessing EOSDIS indicates that services need to serve heterogeneous communities, and services that allow users to access EOSDIS through other systems need to be provided.

NOAA, CIESIN, and other organizations in the U.S., Europe, and Japan will have somewhat similar environmental or Earth Science data and information systems in the 1998-2003 timeframe. The clientele that these systems are being designed to serve will have the option of entering EOSDIS via their own system, or entering EOSDIS directly. In general we expect that users would enter EOSDIS directly, unless their sponsoring organization, e.g., NOAA, CIESIN, DoI, Japan (NASDA/MITI), Europe (ESA, EC) paid (or made other arrangements for) whatever fees NASA would charge for EOSDIS data and services. In these cases they would most likely go through their sponsoring organization's data system.

5.1.3 Access Paths

While access mode describes how the user is interfacing with the system, access path defines the path a user takes from system access to the data server. This information coupled with demographics provides sizing information for various system components. The relative use of the various paths that users take in accessing the EOSDIS data server is dependent upon the degree to which users in each sector are:

- familiar with EOSDIS data sets,
- familiar with the location of desired data sets,
- receiving their data through standing order,
- familiar with EOSDIS services,
- likely to search multiple datasets simultaneously crossing DAACs and/or SCFs,
- exploring data sets and results outside their normal discipline.

The diagram below (Figure 5.1) illustrates the various pathways to the data server including the use of intermediary services (Advertising Service, Distributed Information Manager, Local Information Manager) to assist in locating data products. It is expected that as a user's familiarity with the system increases, his or her pathway through EOSDIS will change to direct access to data servers.

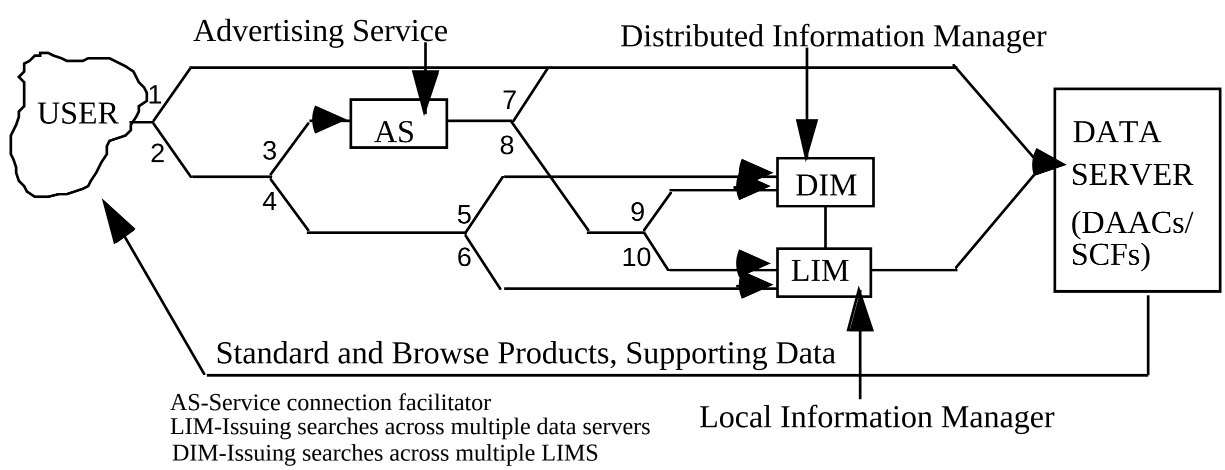


Figure 5.1 Access paths from the user to the data server

5.2 Science User Community

5.2.1 Frequency of System Access

To estimate the frequency of system access by the science user community, a classification was made based upon the number of accesses: yearly = 1-2, quarterly = 3-11, monthly = 12-24, weekly = 25-100, and daily = 100-250. Based upon the insights gained from the science scenarios, the number of users associated with this classification were summed in order to estimate the frequency of user accesses and the number of initial accesses/day to various service components averaged over a year. This was coupled with the demographic estimates and the accesses were mapped into the frequency classification, in order to determine how often the entire science user community would access the system. Table 5-1 summarizes these results. For the maximum user demographics there will be 47 accesses $user^{-1} year^{-1}$ on average.

Table 5-1. Frequency of System Access for Science Users

	Minimum	Maximum	Percentage
Yearly (1-2)	800	1500	13%
Quarterly (3-11)	1850	3500	30%
Monthly (12-24)	850	1600	14%
Weekly (25-100)	1800	3400	29%
Daily (100-250)	800	1600	14%
Note: Numbers in () indicate number of accesses/year			

5.2.2 Modes of System Access

Entry into EOSDIS will be gained in one of two ways: directly or through another data and information system. The following sections describe the methodologies used to estimate system access directly (Section 5.2.2.1) and via other systems (Section 5.2.2.2).

5.2.2.1 Direct Access

The implicit analysis performed when creating the user scenario matrix was used to identify modes of system access. The demographics associated with a matrix cell for each category were added together to get the total number of users and percentage shown in Table 5-2. The modes of access are divided into three major categories: telephone interface only, electronic interface, and machine-to-machine interface. The number and percentage of users that would use each of these access methods is estimated based upon information contained in the science scenarios and in user questionnaires.

The electronic category is further divided into four subcategories: 1.)Standing Orders are received by users desiring regular automatic delivery of data; 2.)Browsers are users that browse the data and do not really do any other analysis on the system; 3.) Remote File Access (RFA) users do some form of analysis on the EOSDIS system using ECS processing resources; and 4.) Data Producers are those users that are using data to produce different and higher-level data products that will be managed by ECS. These subcategories are not mutually exclusive. For example, a Data Producer may be receiving the data he needs to produce his or her product via a standing order.

Table 5-2. Modes of System Access by Science Users (U.S.)

Method	Percentage	Number(Users)
Telephone Interface Only	1.5%	90-170
Electronic	91.5%	5,600-10,600
Standing Orders		1,400-2,700
Browsers		2,00-3,800
Remote File Access (RFA)		2,800-5,300
Data Producers		800-1,500
Machine-to-Machine	7%	430-810
Total	**100%**	**6,100-11,600**

5.2.2.2 Access via Other Information Systems

Estimates of the number of Science users entering EOSDIS directly and through other government data systems are given in Table 5-4. These estimates were based upon discussions with individuals involved with the development and operations of data and information systems in DoI, NOAA, and DoE, as well as prior experience in working with people associated with European and Japanese data systems.

Table 5-4. Number of Science Users Accessing Directly and Through Other Systems

United States	
	Numbers of Users
Direct Access	5,900-11,000
Access Through Other Data Systems:	
NOAA	100-300
CIESIN	?
Other	100-300
Total	**6,100-11,600**
From Other Countries	
Direct Access	2,900-4,500
Access Through Other Data Systems (e.g. Europe, Japan)	1,400-2,000
Total	**4,300-6,500**

5.2.3 Access paths

To estimate the number of first accesses to different service components (Data Servers, Advertising Service, Distributed Information Manager, Local Information Manager) an analysis of the scenarios was conducted to determine how many accesses went to the DSs, AS, DIM and LIMs. Each scenario was analyzed to determine how many accesses to the system would be made to complete the scenario. Knowing the definition of the AS, DIM, LIM, and DS, an educated judgement was made as to which of these services was being utilized first in each access to the system. Upon completion of the scenario analysis, the service accesses were summed by type to get the percentages presented below for the data server access routes presented in Figure 5.1. These first accesses to the system were summed to obtain the total number of accesses. Table 5-3 summarizes the results.

Table 5-3 Fraction of science users accessing EOSDIS data servers via 10 access paths

Path Number	Fraction of Users	Comment
1	.867	86.7% of users will access data servers directly
2	.133	
3	.08	8% of users will make use of the Advertising service
4	.053	
5	.016	1.8% of users, #5 + #9, will make use of the DIM
9	.002	
6	.037	4.5% of users, #6 + #10, will make use of the LIM
10	.008	
7	.07	
8	.01	

The majority of the Science community will access data servers directly. The implication of these results for the ECS developers is that attention to services allowing direct use of data servers by science users should be provided.

5.3 Non-Science User Community

5.3.1 Frequency of System Access

For each sector of the non-science community, an estimate was made of the yearly demand for each of the data products (over 250 in number), including the demand for browse-only products. This estimate was based on a questionnaire of federal agencies, discussions with state organizations, and interviews with educators, together with the derived results regarding the size of the potential user communities. Given these estimates and the total number of users in each sector (Section 4.5), the frequency and number of accesses are calculated and presented in Table

5-5. This analysis yields an upper limit of the total number of accesses of 640,000/year or an average of 3.2 accesses user^{-1} year^{-1} for the non-science community.

Table 5-5. Frequency of Access by Non-Science Users

	Minimum	Maximum	Percentage
Yearly (1-2)	65,200	186,200	93%
Quarterly (3-11)	3,700	10,600	5%
Monthly (12-24)	300	800	<1%
Weekly (25-100)	500	1,400	<1%
Daily (100-250)	300	1,000	<1%
Note: Numbers in () indicate number of accesses/year			

5.3.2 Modes of System Access

Estimates of the number of Non-science users entering EOSDIS directly and through other government data systems are given in Table 5-6. As was the case for the Science user community, these estimates were based upon discussions with individuals involved with the development and operations of data and information systems in DoI, NOAA, and DoE, as well as prior experience in working with people associated with European and Japanese data systems.

Table 5-6 Non-Science Users Accessing Directly and Through Other Systems

United States Only (assumes NASA will serve these communities)	
Direct Access	69,000-198,000
Access Through other data systems (NOAA, DoI, Other)	1,000 - 2,000
Total	**70,000-200,000**

5.3.3 Access Paths

For each sector of the non-science community an estimate was made, based upon the experience of the contributors to this White Paper in working with these communities, as to the percent of users that would follow each of the ten paths identified in Figure 5-1. When these sector estimates are aggregated to the total non-science community, the percentages shown in Table 5-7 result:

The primary users of the Advertising Service, the DIM and the LIM will be the non-science communities and scientists who are seeking data in areas other than in their normal discipline.

Therefore these services need to be designed so as to communicate with a very large and diverse user community.

Because of the potentially large number of Non-science users, attention must be paid to the management of resources to allocate ECS services on a priority basis and to encouraging other value-added service providers to serve the non-science communities. This also implies that EOSDIS must be able to accommodate interaction with other service providers.

Table 5-7. Fraction of non-science users accessing EOSDIS data servers via 10 access paths

Path Number	Fraction of Users	Comment
1	.04	4% of users will access data servers directly
2	.96	
3	.86	86% of users will make use of the Advertising service
4	.10	
5	.05	9% of users, #5+#9, will make use of the DIM
9	.04	
6	.05	9% of users, #6+#10, will make use of the LIM
10	.04	
7	.78	
8	.08	

6.0 Data Access Characteristics

6.1 Introduction

Several factors related to access and distribution of data are important to consider. The volume of data accessed at each DAAC by each user community, the mode of data distribution (i.e., physical media vs. electronic) to each community, and the fraction of orders which are filled via standing order vs. ad hoc requests will affect the size of the system components as well as I/O characteristics (i.e., time between a request and when the data is staged to the data server) and communication bandwidth. In addition, the ratio of the data volume staged to the data volume distributed to the users indicates the amount of subsetting required to fulfill user requests.

The important factors in the determination of the volume of data to be accessed and distributed are the relative interest that the various user communities will have for the available data products, and the size of the associated communities. Different methodologies are used to estimate the size of the projected demand from the science and non-science communities and these are described below for two main user categories (General Science and Non-science) in the sections that follow.

6.2 Interest in Data by Pyramid Layer

The Data Pyramid was developed as a tool to conceptually organize the many types of data that the EOSDIS will be responsible for managing. The upper layers of the pyramid contain information to assist users in using the EOS data and the lower layers are made up of the different data product levels (0, 1A, 1B, 2, 3, and 4).

The level of interest in each layer of the data pyramid provides ECS developers with information regarding the expected load on the data server at each DAAC. This is due to the fact that it is already known where each product will be archived. Thus, level of interest combined with demographics and data volumes required by users will indicate the loads generated at each DAAC by the overall user community. Different methodologies are employed for the science and non-science communities; these are described in the sections that follow.

6.2.1 Science Community

The relative interest in the standard data products from the science community was determined, in part, by the relative discipline focus of the EOS and General Science communities. One key input to this analysis is a table of the Relative Product Access Frequencies (RPAFs); the entries in this table provide the frequency with which a granule of one product is expected to be accessed relative to granules of another product. It is a necessary input to the overall modeling effort in that the models (in particular the Static Model) need to address issues related to user access of products on a DAAC-by-DAAC basis. This data was also used to extend the narrow product scope of user scenarios to the full range of available products.

6.2.1.1 Interpretation of Relative Product Access Frequencies (RPAFs)

The RPAFs are, as described above, a measure of the frequency with which granules of one product are accessed relative to granules of another. The RPAFs are interpreted as in the following example: if the RPAF for MOD09 (Surface Reflectance) is 2.044×10^{-2} and for SEA04 (K 490 product, mapped) is 2.913×10^{-5}, then by dividing the RPAF for MOD09 by the RPAF for SEA04, one estimates that accesses of granules of MOD09 by science users will be approximately 700 times more frequent than for granules of SEA04.

6.2.1.2 Procedure for Determining RPAFs

The RPAF is based on the fact that each product can be classified according to the science discipline which that product serves, and that members of a discipline will predominantly use discipline-specific data. A coarse product classification scheme is employed which contains 6 classes: Atmospheric, Land, Ocean, Cryospheric, General, and Miscellaneous. General means there is no particular discipline associated with a product (e.g., Level 1 data), and Miscellaneous means that the product is not specifically useful to any discipline (e.g., Level 0 data and book-keeping data). These are 'roll-ups' of the classes in the Data Model's 'data type collections' except for the General and Miscellaneous categories Here, however, products can be associated with multiple disciplines because the parameters they contain are associated with different disciplines.

Having classified the products according to discipline, a parallel approach is taken with respect to science user classes: Atmosphere, Land, Oceans, Cryosphere, and Interdisciplinary. These science user classes affect the RPAF in two ways: first, via the relative size of each class (Table 6-1), and second, via each class' relative interest in products of each product class (Table 6-2). The rationale for all assumptions is given in Section 6.2.1.3.

An estimate of relative interest in product levels (Levels 0, 1A, 1B, 2, 3, and 4) must be applied (Table 6-3), and finally, the relative number of granules for each product is applied.

Table 6-1. Relative sizes of User Disciplines

Discipline	Atmos	Land	Ocean	Cryo	Interdisc
Relative Size	0.50	0.27	0.18	0.01	0.04

Formally, all of this information is combined as follows:

$$\text{RPAF(PROD\_i)} = \left\{ \sum_{ud} \left[F(ud)\ I(ud,pd_i) \right] G_i\ I(pl_i) \right\} / \text{NORM},$$

where F(ud) is the fraction of users in discipline *ud*, $I(ud,pd_i)$ is the relative interest each user discipline *ud* has in PROD_i's product discipline *pd*, G_i is the number of granules of PROD_i, and I(pl_i) is the relative interest in PROD_i's product level *pl*.. NORM is chosen such that the sum of the RPAFs is 1.

Table 6-2. User Discipline (columns) interest in Product Discipline (rows).

Relative interest is weight normalized to one for each User Discipline.

	Atmos		**Land**		**Ocean**		**Cryo**		**Interdisc**	
	Weight	Rel. Int.	Weight	Rel. Int.	Weight	Rel. Int.	Weight	Rel. Int.	Weight	Rel. Int.
A	400.00	0.797	143.00	0.283	28.00	0.262	147.00	0.291	59.00	0.518
L	40.00	0.080	286.00	0.565	6.00	0.056	29.00	0.057	25.00	0.219
O	40.00	0.080	14.00	0.028	56.00	0.523	29.00	0.057	19.00	0.167
C	16.00	0.032	57.00	0.113	11.00	0.103	294.00	0.582	6.00	0.053
G	5.00	0.010	5.00	0.010	5.00	0.047	5.00	0.010	4.00	0.035
M	1.00	0.002	1.00	0.002	1.00	0.009	1.00	0.002	1.00	0.009

Table 6-3. Relative interest in Product Levels

Product Level	**0**	**1a**	**1b**	**2**	**3**	**4**
Weight	1.00	10.00	100.00	100.00	100.00	100.00
Rel. Int.	0.002	0.024	0.243	0.243	0.243	0.243

6.2.1.3 Basis of Estimates and Assumptions

6.2.1.3.1 Relative User Discipline Sizes

The relative size of each user discipline is given in Table **6.**. These numbers are based on:

1) Miller's science literature survey (see Section 4.3) which showed that the relationship between the Atmosphere, Land + "Other", and Oceans disciplines is approximately:

Atmos:Land+Other:Ocean::50:32:18.

2) Conversations with Miller in which he stated his impression from that survey is that the Cryosphere Discipline is approximately 1% of the total community, or:

Cryosphere:Total::1:100.

3) Theobald's observation that approximately 4% of the EOS investigators are Interdisciplinary, or:

Investigators :Total::1:25.

Thus, for every 100 science users, 50 users will be a member of the Atmospheric discipline, 18 will be in the Oceans discipline, 4 will be Interdisciplinary investigators, 1 will be in the Cryosphere category, and the balance (27) will be from the Land Science discipline; put another way, the ratios between sizes of these 5 science disciplines are:

Atmos:Land:Oceans:Cryo:Interdisc::50:27:18:1:4.

6.2.1.3.2 Relative Product Level Interest

We assume that the interest in product levels are as follows: Levels 1B, 2, 3, and 4 are 10 times as likely to be accessed as Level 1A data, and 100 times as likely to be accessed as Level 0 data

In that these numbers are supposed to describe the General Earth Science community , it is unlikely that they will use Level 0 or Level 1A data, somewhat likely that they will use Level 1B, and more likely that they will use higher level products. However, because we want a conservatively high estimate of demand, these ratios are probably lower than they should be. Based on these assumptions, the relative interest can be expressed as:

Level1B:Level2:Level3:Level4:Level1A:Level0::100:100:100:100:10:1.

6.2.1.3.3 Relative User Discipline Interest in Product Discipline

For those members of a particular User Discipline that have varying (non-zero) interest in products from outside their discipline, the following assumptions were made to calculate an RPAF. Note that the values in the descriptions below do not exactly match those in Table 6-2**;** this is due to the fact that the values in the table are calculated from a formula whereas the values in the descriptions are order of magnitude approximations.

Atmospheric discipline users Atmospheric discipline users will have an interest in Land and Ocean data because they may need to eliminate the effects of Land and Ocean processes in their atmospheric data. Because these Land and Ocean processes generally occur on longer time scales than atmospheric data, we assume that their interest in these types of products is in the ratio Atmos:Land:Ocean::10:1:1 (approximately daily:weekly). Similarly, for atmospheric user interest in Cryospheric data, we assume an even longer time scale: Atmos:Cryo::25:1

(approximately daily:monthly). We assume that 1% of the users will be interested in General data and that interest in General data is five times greater than for Miscellaneous data, resulting in the ratios General:Total::1:100 and General:Misc::5:1. Thus, we have the relationship:

Atmospheric: Land:Ocean:Cryo:Gen:Misc::400:40:40:16:5:1,

or, for every 500 products ordered by Atmospheric discipline users, 400 will be Atmospheric, 40 will be Land, 40 will be Ocean, 16 will be Cryospheric, 5 will be General, and 1 will be Miscellaneous.

Land discipline users We assume that atmospheric products remain a large factor in Land science analyses and that every other Land product will need corresponding atmospheric data, or Land:Atmos::2:1. We assume that use of Ocean products will be a minor component - Land:Ocean::20:1, but that Cryospheric products (which include snow cover, etc.) will be accessed in the ratio Land:Cryo::5:1. For General and Miscellaneous products we assume the same ratio as for Atmospheric users: Total:General:Misc::500:5:1. Thus, for land discipline users, the relative interest in product disciplines is:

Land:Atmos:Cryo:Ocean:General:Misc::300:150:60:20:5:1,

or, for every 500 products accessed by Land users, approximately 300 will be Land products, 150 will be Atmospheric, 60 will be Cryospheric, 20 will be Ocean, 5 will be General, and 1 will be Miscellaneous.

Oceans discipline users We make the same assumption for these users regarding Atmospheric data as was made for the Land discipline users: Ocean:Atmos::2:1. We assume that Ocean discipline users will need Land data as often as atmospheric discipline users will, or Ocean:Land::10:1, and the same relative need for Cryospheric data (which includes sea ice, etc.): Ocean:Cryo::5:1. We assume a slightly higher need for General data by the Oceans community than for the others, as this contains Level1b products: Ocean:General::11:1; and the same assumption for Miscellaneous products: General:Misc::5:1. Thus the relative interest ratios for the Oceans community are:

Ocean:Atmos:Land:Cryo:Gen:Misc::100:50:10:11:5:1

Cryospheric discipline users We again assume that Cryospheric discipline users will need atmospheric data for every other Cryospheric product they use; thus the ratio is Cryo:Atmos::2:1. We make the assumption for both Land and Ocean data that they are needed 10% as often as explicit Cryospheric data, resulting in the relationship Cryo:Land:Ocean::10:1:1, and we make the usual assumptions regarding General and Miscellaneous data: Total:General:Misc::500:5:1. The overall relationship for Cryospheric discipline users is then

Cryo:Atmos: Land: Ocean:Gen: Misc::300:150:30:30:5:1.

Interdisciplinary users For Interdisciplinary users, we rely on data from the SPSO regarding EOS Interdisciplinary Science Investigation (IDS) data product needs. Based on an analysis of the data needs of the IDS users, we have the relationship:

Atmos:Land:Ocean:Cryo:Gen:Misc::59:25:19:6:4:1

6.2.2 Non-science Community

The relative interest in the data pyramid layers of the non-science community was based upon the returned questionnaires from the Federal agencies, discussions with state organizations, and interviews with educators. Judgements were made as to the interest of each of the non-science sectors in each of the more than 250 standard data products, the geographic scale of interest for each product, and the number of times per year users would access each product and geographic scale. These estimates were then used to determine the number of user accesses for each layer of the data pyramid.

6.2.3 Results and Discussion

Based upon the methodologies described above, Figure 6.1 depicts the relative interest of each data pyramid layer of each user community. Relative interest is expressed as the per cent of user accesses that act upon each pyramid layer.

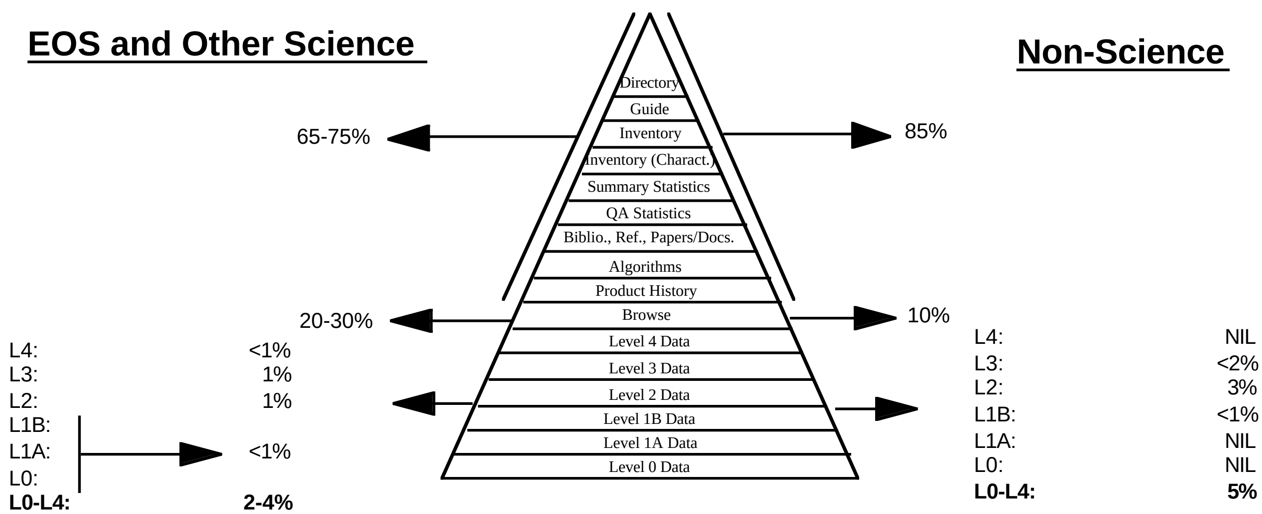


Figure 6.1 Per cent of user accesses operating on each layer of the data pyramid

As expected, Figure 6.1 suggests that the Science community will spend more time with the actual data than will the non-science community. The non-science users would make more use of the upper layers of the pyramid to obtain descriptive information about the data. In both cases, the results confirm the need for rapid access to the descriptive information about the data products. Also, since the access profiles across DAACs are rather heterogeneous, different design solutions are called for with different resource requirements.

The reason for the low projected demand for the Level 4 products by the non-science community is due to the fact that very few such products are currently included in the Standard Data Product lists. As more descriptive information regarding Level 4 products becomes available, it is expected that the projected demand will increase significantly. Also, while the percentages of accesses of browse products represents only 10% of the total accesses by this community, the number of browse products to be accessed per year could be in excess of 11,000,000 with the primary demand coming from the educational community. However, there is currently very little information available regarding the characterization of browse products to be produced, by whom, and when. Once this information is available, a more accurate assessment of the demand for browse products will be possible.

6.3 Data Volumes Staged and Distributed

Estimating the data volume that users are expected to pull from the data archive is important because it affects storage loading, I/O performance, and computing resources. This information can be used to size system components and to understand the response time between a user's request and the fulfillment of that request. Comparing the data volume pulled from the archive to the volume distributed to the user indicates the amount of subsetting that the system must perform to satisfy the user's request.

6.3.1 General Assumptions

The volumes estimated are for the 1999-2003 time frame. It is assumed that by this time, direct parameter-level access is possible, both out of the archive and from the processing stream. Products, product sizes, and granule definitions are consistent with the SDR product baseline as of May 10, 1994. While estimates were made for the projected demand for Landsat and SAR data products, time did not permit the inclusion of this data at SDR. Since there is currently no adequate definition of browse products that will be available, the browse volumes are not included. However, there appears to be a strong potential demand for browse products, especially for the General Science and Non-science communities.

The results are also based upon a 365 day per year system operation with a 250 day per year user work year. There is no latency assumption in “user retrieval” estimates (other than 365 vs. 250). Data volumes due to standing orders and ad-hoc requests were combined, as well as volumes from electronic and media transfer. The availability of a “smart” subscription service was assumed where initial subscription enrollment includes user specification of desired geographic regions and parameters in which only the data whose granule boundaries satisfy those specifications are staged/distributed.

6.3.2 Results and Discussion

To obtain an estimate of the volume of data to be pulled by the user communities, summaries were prepared of the total demand in a year for each standard data product, for each geographical scale (i.e., $1\text{x } 10^2 \text{ km}^2$, $1\text{x}10^3 \text{ km}^2$, $5\text{x}10^5 \text{ km}^2$ and $1\text{x}10^8 \text{ km}^2$) and for each sector of the user community based upon the results of the methodologies described in Section 6.2 of this document. These summaries were used as input to the Static Model (science users) and the Volumetric Model (non-science users). Knowing the size of each data granule and the subsetting required, the volume, both staged and distributed to the user, was calculated.

Table 6-4. Total Data Volumes Staged and Distributed

TOTALS	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	1,906,954	4,662.746	1,196,523	2,626,236
GSFC	2,626,803	5,964,737	2,259,538	4,507,033
JPL	TBD	TBD	TBD	TBD
LaRC	1,176,414	3,152,429	460,055	1,206,762
MSFC	81,907	196,092	15,970	48,210
NSIDC	146,932	417,348	58,952	162,134
ORNL	TBD	TBD	TBD	TBD
TOTAL	**5,939,010**	**14,393,351**	**3,991,038**	**8,550,374**

Inspection of Table 6-4 shows that the range of estimates spans a factor of 2 to 3 (does not include Landsat data). Relative to the production volume (2.1 TB/day), the volume staged in the data server due to user pull is greater by a factor of 2.8 (min) and 6.7 (max). The volume distributed to all users is greater than the production volume by a factor of 1.9 (min) and 4.0 (max). The subsetting ratio is 2-3:1 (on average, after correcting for 250 vs. 365 day years). This ratio is substantially higher at some DAACs and for some user groups: 10-1000:1 (See Figure 6.2). Smaller granules will reduce the amount of subsetting required and the volume staged. It is also important to note that the volume/request (~2 GB/request, not shown in table) is consistent with current Goddard DAAC experience with "EOS-like" data. Calculations were based upon the granule sizes contained in the SPSO database which has not been verified by the instrument teams.

The data volumes for the IDS investigators Table 6-5a are worst case estimates provided by NASA. Therefore, the numbers shown in Figure 6.2 and in Tables 6-5a and 6-4 are believed to overstate the projected volume distributed. Figure 6.2 gives the resulting bounding estimates for the total user pull requirements for data staged in the data servers, subsetting amount (implicit), and distribution.

Table 6-5a. EOS IDS and Instrument Investigators' Data Volumes Staged and Distributed

EOS IDS +Instrument	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	522,826	980,713	763,849	1,432,821
GSFC	1,042,165	1,911,298	1,522,603	2,792,407
JPL	TBD	TBD	TBD	TBD
LaRC	91,523	151,432	133,715	221,242
MSFC	170	338	248	494
NSIDC	6,723	13,318	9,822	19,458
ORNL	TBD	TBD	TBD	TBD
TOTAL	**1,663,406**	**3,057,099**	**2,430,236**	**4,466,422**

Table 6-5b. General Science Data Volumes Staged and Distributed

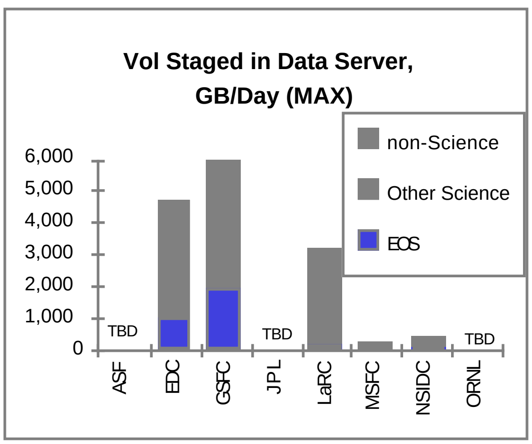
Other (General) Science	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	984,378	2,989,236	355,482	1,079,483
GSFC	1,076,938	3,270,310	388,908	1,180,986
JPL	TBD	TBD	TBD	TBD
LaRC	894,986	2,717,781	323,200	981,454
MSFC	43,484	132,046	15,703	47,685
NSIDC	124,101	376,854	44,816	136,091
ORNL	TBD	TBD	TBD	TBD
TOTAL	**3,123,888**	**9,486,228**	**1,128,109**	**3,425,699**

Table 6-5c.   Non-Science Data Volumes Staged and Distributed

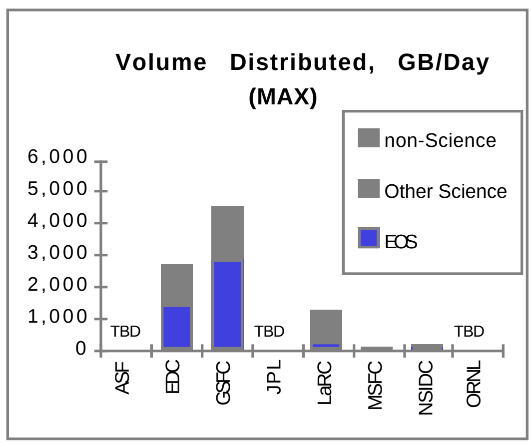
Non-Science	Volume Staged in Data Server (MB/day, 365 days/year op'n)		Volume Distributed to Users (MB/day, 250 days/year op'n)	
DAAC	**MIN**	**MAX**	**MIN**	**MAX**
ASF	TBD	TBD	TBD	TBD
EDC*	399,750	692,797	77,193	113,932
GSFC	507,700	783,128	348,027	533,640
JPL	TBD	TBD	TBD	TBD
LaRC	189,905	283,216	3,139	4,066
MSFC	38,253	63,708	19	31
NSIDC	16,108	27,176	4,315	6,585
ORNL	TBD	TBD	TBD	TBD
TOTAL	**1,151,717**	**1,850,024**	**432,994**	**658,254**

6.3.2.1 Data Volumes Staged and Distributed by DAAC

Figure 6.2 graphically shows the volume of data staged to the data servers in order to fulfill the requests of EOS science users, general science users, and non-science users. Also shown is the volume of data distributed to each of these communities. Comparison of the two volumes yields an estimate of the amount of data subsetting that each user community will require.



Total = 14,393 GB/day
(5,939 GB/day - MIN)



Total = 8,550 GB/day
(3,991 GB/day - MIN)

Figure 6.2.   Data volume staged and distributed to EOS science users, general science users, and non-science users per day (average per day based on a 250-day year)

6.3.2.2 Data Volumes Distributed by Pyramid Layer

Figure 6-4 depicts the projected volume of data that will be distributed to users from different parts of the data pyramid. The data volumes from the upper layers could not be broken out by individual layer but since the layers contain similar types of data, they are grouped together and only one total volume is provided. It is apparent that most of the volume of data distributed is from the lower levels of the data pyramid.

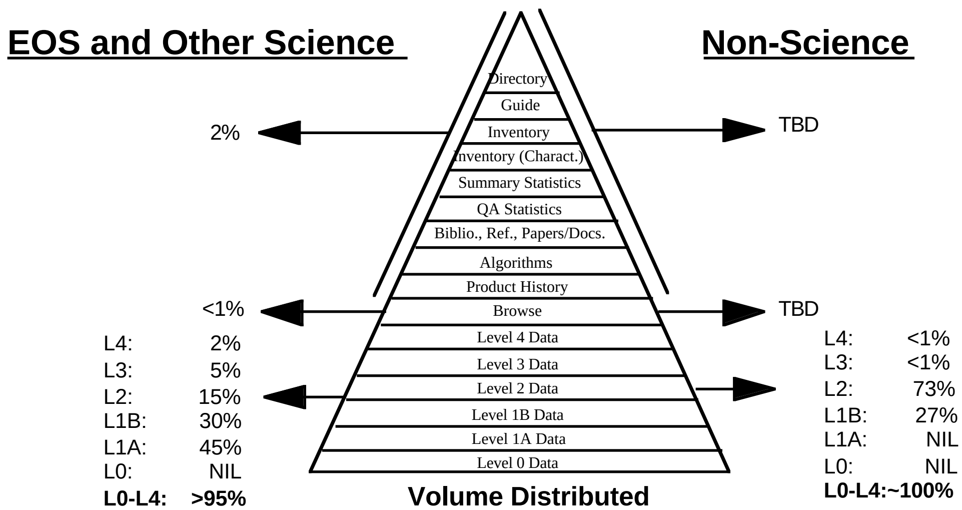


Figure 6.3. Per cent of data volume accessed by general science users and non-science users as a function of data pyramid layer.

Appendix A: December, 1993 User Matrix and Class Definitions

This appendix contains the user scenario matrix that was in use in the December, 1993 timeframe. Also included are the definitions of each user class that correspond to the classes in the user matrix.

December, 1993 User Scenario Matrix

	General Info Searches	Reviews	Theoretical Studies	Case Studies	Field Studies	Climatologies/Global
Intermediary to Education or Policy Community (e.g., CIESIN, S4 proposals	Intermediary to Dept. of Education; high-level summary of meteorological data for grades K-12 *Bill Emery* 1	Lawyer hires intermediary; summary of snowfalls for lawsuit against a ski company *Edward Calvin* *Tyahla/Theobald* 2	Sociologist; hypothesis- people of means live upwind of industry in urban areas. *Dave Walker* *Tyahla/Theobald* 3	Writer for McGraw-Hill needs to prepare a text demonstrating EOSDIS via progressively complicated examples 4	Intermediary under contract to Dept. of Ed. prepares science lesson plans for Internet-wide distrib. 5	Sociologist-"people/park conflict"- 25 large game reserves in sub-Saharan Africa. *Michael Garstang* *Tyahla/Theobald* 6
Traditional User contacting EOSDIS directly	High School Teacher; wants students to get radiance data to correlate with properties of river water samples *Donald Foss* *Lori Tyahla*	Virginia Coast Reserve Long-Term Ecological Res. Prog. - mapping and tracking vegetation dynamics *Raymond Deuser* *Tyahla/Theobald* 8	Test ecological theory regarding vegetation competition in grasslands across the central U.S. *Don Strebel* *Celeste Jarvis* 9	Insurance Co. Rep.; wants geographical extent of Mississippi River Flood to verify claims *Bill Kennedy* *John Daucsavage* 10	Cryosphere; researcher using surface reflectance to determine age of ice surface on land *Chris Shuman* *Celeste Jarvis* 11	Intn'l Monetary Fund; wants data to verify credit worthiness of multi-billion dollar loan for irrigation project 12
Character text user	News reporter; wants before and after photos of Mississippi River flood area *Bill Kennedy* *John Daucsavage* 13	Undergrad. Student in intro. to Remote Sensing needs to research what instruments/data sets are compatible with senior thesis *Jan Poston* *Lori Tyahla* 14	NOAA researcher studying seasonal and diurnal variation in regional lightning distribution *Raul Lopez* *Lori Tyahla* 15	Forest Ranger preparing a report for a Department of Interior Policy Maker needs pre- and post- forest fire data to assess recovery *Donald Ohlen* *John Daucsavage* 16	An oil company needs regional geological and vegetative data to determine best drilling sites. *Bill Kennedy* *John Daucsavage* 17	Political Science Professor at a small college wants to correlate NDVI data with global population and GNP data *Jeff Eidenshink* *John Daucsavage* 18
Data Consumer (Moderate Access)	A local government near LA wants daily ocean color data delivered once/month (algal growth) *Carolyn Whitaker* 19	Earth Science Researcher wishes to access electronic journal *Jeff Dozier* *Lori Tyahla* 20	NSIDC Scenario #3 Snow depth and Extent; Polar Jet Stream *John Walsh* *Khalsa/Kaminski* 21	MSFC Scenario #2 Global wind field detection; aerosol backscatter-case study oriented *Dave Emmitt* *Theobald/Tyahla* 22	ISI Global Water Cycle;, includes model verification through field studies; *Eric Barron* *Lori Tyahla* 23	NSIDC Scenario #1 Surface and top-of-atmosphere radiative fluxes over sea ice during summer (2 yrs.) *Jeff Key* *Khalsa/ Kaminski* 24
Data Browser (Frequent Access)	Research Librarian *Cristina Sharretts* *Tyahla/Theobald* 25	Investigation of algorithms involving a wide range of EOS instruments which will provide detection, tracking, and warning of volcanic events and ejectamenta. 26	Earth Science Community User; e.g., University Prof., Radiation Budget *Barkstrom (CERES)* *Haldun Direskinelli* 27	Instrument Support Terminal User; e.g., ASTER Team Member *Bob Hekl* *Tyahla/Theobald* 28	Use of Cryospheric System to Monitor Global Change in Canada; *Rejean Simard* *Lori Tyahla* 29A Arctic Ice pack response to weather *John Heinrichs* *Celeste Jarvis* 29B	Changes in Biogeochemical Cycles; *Berrien Moore, III* *Mike Theobald* 30
Analytical User (Frequent Access)	31	*H. Grant Goodell* *Tanya Furman* *Tyahla/Theobald*	Stratospheric chemisry and dynamics *Leslie Lait* *Mike Theobald* 33	Detection and classification of transparent cirrus clouds. *Dan Baldwin* *Tyahla/Theobald* 34	Interdisciplinary Ocean/Atmosphere Field Campaign (a la TOGA-COARE *Jim Wang & David Short* *A. K. Sharma* 35	Climate, Erosion, and Tectonics in Andes and other mountain systems; *Bryan Isacks* *Theobald/Tyahla* 36
Production User (Frequent Access)	37		MSFC Scenario #1 Validation of passive microwave algorithm for precipitation retrieval *Michael Goodman* *Danny Hardin* 39	Commercial User; value-added products *John Daucsavage* 40	Interdisciplinary Investigation of Clouds and Earth's Radiant Energy System; *Bruce Wielicki* *Mike Theobald* 41	GCM Modeler ; *Jim Stobie* *Celeste Jarvis* 42A --------------------------------- EOS Instrument Investigator; e.g., MODIS, Ocean Color *Mark Abbott* *Celeste Jarvis* 42B
Advanced Technology User	43	44	Intn'l Interdisciplinary PI; e.g., will event recognition software work on L4 data to flag a particular event? *Mouginis-Mark* *Lori Tyahla* 45	Development of Automated Snow Mapping Procedure (Sequoia 2000 Scenario) *Walter Rosenthal* *Lori Tyahla* 46	Calibration/Validation of MODIS Ocean Products *Bob Evans* *T heobald & Tyahla* 47	AIRS Team 48

System Access Pattern Definitions (rows of matrix):

Person-to-Person: This group of users prefers to search for required data by phoning the science support staff at one of the DAACs. They will describe their current research and inquire about suitable data (or they may already know what data they require, but prefer to order it via phone, FAX, letter, or e-mail). They generally are supported by an institutional computer facility and programming staff. At most, this user may wish to access the data via a CDROM.

Character Text User: This user probably has used text type data listings on VT100 type terminals throughout his or her career and sees no reason to change solely for the purpose of ordering required data. He or she is likely to be used to receiving data on 9-track tapes and having research assistants write software to ingest and analyze the data via an institutional computing department.

Data Consumer: A user in this group tends to desire routine shipment of a compiled and massaged set of data with built-in software for read and display. This type of data consumer would probably access the system with moderate frequency and would prefer the data on CDROM and might use a souped-up Macintosh to look at it.

Data Browser: Although every user will probably browse data before ordering it, the users in this category are required to browse frequently in order to perform their jobs. For instance, a Field Support Terminal User must browse data in the field in near real-time as the field investigation is taking place in order to properly control the experiment. An Instrument Support Terminal User will browse data to monitor instrument health and safety. Others in this group include those users who may need to browse a large amount of data in order to find existing data that may affect their current theoretical investigation. This group also includes the individual who enjoys browsing data looking for interesting and possibly previously unnoticed coincidences.

Analytical User: Analytical users are expected to access the ECS frequently for a variety of purposes. This group includes calibration experts and algorithm/model validators who may not require large quantities of data per access, but may access the ECS quite frequently. In addition, this group also includes researchers who are familiar with the data they require and may not browse it before ordering a large amount of it.

Production User: Production Users are also expected to access ECS frequently. Users in this group are generally producing higher level data products from raw data or lower level products. Some of the products are standard products produced by the ECS and others will be produced by algorithms developed outside the ECS. This group also includes the Data Maintainer who decides whether or not data needs to be reprocessed, as well as Modelers who desire Level 2 or 3 products routinely piped directly to the model over the network.

Advanced Technology

Users: These users possess workstations and software which utilize the very latest technology. They require highly specialized communication services such as video, voice, animation, and visualization concurrently scientist to scientist. These users would like to specify search criteria, computation steps, processes for event recognition, and links between the servers and home analysis package in natural language using terms specific to research.

Data Access Pattern Definitions (columns of matrix):

General Information

Searches: This category is meant to represent data accessed by non-science users such as Policy Makers, Commercial Users, and interested members of the general public. Data accessed in this category is primarily high level data (but possibly Level 2 data) and the data volume per access is relatively low.

Reviews: The type of data accessed for review purposes will probably be bibliographic references to published papers and documentation regarding the data sets. The volume of data accessed this way is still relatively small, but is probably larger than in a general information search.

Theoretical Studies: Theoretical studies generally report on mathematical relationships. Data access for theoretical studies tends to be infrequent and in small amounts (Barkstrom, 1991). Some examples are population genetics in ecological systems and light penetration in the ocean.

Case Studies: Data accessed as case study data are data that are associated with a specific event. For example, one might be interested in all available data regarding Hurricane Hugo, or data related to a single woodlot or forest clearing. Also, one might wish to access NDVI data for a particular area after a fire has occurred.

Field Studies: Barkstrom (1991) defines field studies as studies of a particular region with several different instruments over a limited period of time. Users accessing field study data probably intend to intercompare several different kinds of data; thus they will require special methods for colocating data, editing data, and for calibration. The data volumes accessed here tend to be large.

Climatological/Global

Studies: Users accessing data for climatological studies will require the largest volume of data. Some climatological studies include the entire globe, while others do not.

Abbreviations and Acronyms

AGU	American Geophysical Union
AS	Advertising Service
ASF	Alaska SAR Facility
CIESIN	Consortium for International Earth Science Information
CD-ROM	Compact Disk - Read Only Memory
CDRL	Contract Data Requirements List
CSMS	Communications and System Management Segment
DAAC	Distributed Active Archive Center
DIM	Distributed Information Manager
DoD	Department of Defense
DoE	Department of Energy
DoI	Department of Interior
DS	Data Server
EC	European Community
ECS	EOSDIS Core System
EDC	EROS Data Center
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
EPA	Environmental Protection Agency
EROS	Earth Resources Observation System
ESA	European Space Agency
F&PRS	Functional and Performance Requirements Specification
FAX	facsimile
GB	Gigabyte
GByte	Gigabyte
GIS	Geographic Information System
IDS	Interdisciplinary Science
IEEE	Institute of Electrical and Electronics Engineers

JGR	Journal of Geophysical Research
K-12	Kindergarten through 12th Grade
LIM	Local Information Manager
MITI	Ministry of International Trade and Industry
MOD09	MODIS product #09
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency
NLDN	National Lightning Detection Network
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
Op'n	Operation
PDR	Preliminary Design Review
PI	Principal Investigator
R&D	Research and Development
RFA	Remote File Access
RPAF	Relative Product Access Frequency
RRDB	Recommended Requirements Data Base
SAR	Synthetic Aperature Radar
SCF	Science Computing Facility
SDPS	Science Data Processing Segment
SDR	System Design Review
SDS	System Design Specification
SEA04	SeaWiFs product #04
SPSO	Science Processing Support Office
TBD	To Be Determined
USDA	United States Department of Agriculture